

Enhanced metacognition in autism when perceptual decisions rely solely on sensory evidence

Laurina Fazioli¹

laurina.fazioli@hotmail.fr

Bat-Sheva Hadad¹

bhadad@edu.haifa.ac.il

Rachel N. Denison²

rdenison@bu.edu

Amit Yashar¹

amit.yashar@edu.haifa.ac.il

¹The School of Developmental Neurodiversity in Education, Faculty of Education, University of Haifa

²Department of Psychological and Brain Sciences, Boston University

Abstract

Atypical metacognition has been suggested to underlie autistic phenotypes, given its role in social cognition and behavioural flexibility. However, no study has quantitatively assessed metacognitive abilities in autism. Here, we measured meta-uncertainty—the noise corrupting the estimates of one's own decision uncertainty—in autism. In three experiments, autistic and non-autistic participants ($N = 145$) performed orientation categorisation tasks while simultaneously reporting their choice confidence. By independently manipulating each Bayesian component—sensory uncertainty, prior, and reward—and fitting a recently established process model, we assessed metacognitive abilities and their contingency on the Bayesian components while controlling for first-order decisions. Unlike non-autistic participants, autistic participants' meta-uncertainty depended on which decision component was manipulated, and was lower than that of non-autistic participants specifically when decisions were adjusted for sensory uncertainty. These findings reveal that metacognition in autism is not generally reduced but rather enhanced for inferences that rely primarily on sensory information.

Keywords: autism, metacognition, decision-making, Bayesian perception

In acknowledgment of the ongoing discourse regarding terminology for individuals diagnosed with autism, we use "autistic individuals" and "non-autistic individuals" in line with recent conventions.

Introduction

Metacognition—the ability to monitor and evaluate one's own mental states—plays a fundamental role in human cognition. It allows individuals to assess the reliability of their perceptions, guide learning, and adjust behaviour in uncertain environments¹. Failures of metacognition have been implicated across a range of psychiatric and neurodevelopmental conditions^{2,3}. In autism spectrum disorder (ASD)—marked by atypical social cognition, restricted and repetitive behaviour, and altered sensory processing⁴—impaired metacognition could contribute to alterations in both social reasoning and perceptual processing. However, despite intense interest, the nature of metacognition in autism remains unclear. Some accounts suggest a broad reduction in metacognitive ability⁵, while others point to more selective alterations—such as difficulties in metacognitive control⁶ or in tasks involving social cognition⁷. Previous research has primarily relied on the accuracy of confidence reports in memory, cognitive, or perceptual tasks (reviewed in ref.⁵). However, confidence reports reflect not only self-monitoring abilities but also general task performance and biases in confidence decision boundaries^{8–12}, making it difficult to isolate genuine metacognitive differences.

Here, we adopt a computational approach that directly formalises the processes underlying confidence¹³. We draw on Bayesian theories of perception and decision making^{14,15}, according to which perceptual decision-making is formalised as an inference process that combines sensory uncertainty (i.e., likelihood), prior expectations (i.e., internal models), and reward (i.e., cost function) to compute a decision criterion that minimises expected cost^{14,15}. Recent perceptual categorisation studies have shown that, contrary to longstanding views^{16,17}, autistic individuals integrate all Bayesian components in perceptual decision-making in a manner similar to non-autistic individuals^{18,19} (Fazioli et al., in review). Nevertheless, it remains possible that atypicalities in perceptual decision-making arise at the level of metacognitive monitoring in autism.

Across three experiments, autistic and non-autistic participants performed a perceptual categorisation task (first-order decision-making) while reporting their confidence (second-order decision-making). By independently manipulating prior probabilities, reward structures,

and sensory uncertainty, we were able to ask whether—and how—distinct Bayesian components shape metacognitive ability in autism. Using a recently developed process model of metacognition (the ‘CASANDRE’ or ‘confidence as a noisy decision reliability estimate’ model)¹³, which quantifies the noise corrupting internal estimates of uncertainty (hence, meta-uncertainty), we obtained a measure of metacognitive abilities that is independent of task difficulty and confidence bias. Our quantitative approach to metacognition reveals a qualitative divergence between autistic and non-autistic individuals in how Bayesian factors contribute to metacognitive ability. While for non-autistic participants, metacognitive abilities remained stable across experiments, autistics’ meta-uncertainty varied depending on which Bayesian information biased first-order decisions. Specifically, they exhibited enhanced abilities when perceptual decisions relied on sensory information alone, relative to non-autistic participants, and their own performance during integration of prior or reward information.

Results

Participants (52 autistic and 93 non-autistic) categorised the orientation of grating stimuli (first-order task) and reported their confidence (second-order task) on every trial (**Fig. 1a**). In all experiments, to manipulate sensory uncertainty, we varied stimulus contrast across seven values. We manipulated information level regarding orientation category by randomly varying the stimulus orientation on each trial. In Task 1, for each trial, stimulus orientation was drawn from one of the two partially overlapping Gaussian distributions^{20–22}, with means $m_A = -4^\circ$ (Category A) and $m_B = 4^\circ$ (Category B) and standard deviations $s_A = s_B = 5^\circ$ (**Fig. 1b, Task 1**). Participants reported simultaneously on the stimulus category (Category A vs. B) and confidence rating (4-point scale) by pressing one of eight keys, ranging from Category A highly confident to Category B highly confident (**Fig. 1a**). This simultaneous report prevented post-decision bias²³.

In Task 1, we tested how participants made second-order confidence decisions when first-order perceptual decisions integrated prior (Experiment 1) or reward (Experiment 2) information with sensory uncertainty. In Experiment 1, we manipulated priors by explicitly varying category base rate probability across blocks. On a given block of trials, categories could appear with balanced (Category A = 50% and Category B = 50%) or unbalanced base rate probabilities (Category A = 25 % and Category B = 75%, or Category A = 75 % and Category B = 25%). In Experiment 2, we varied the points awarded for correct answers across three blocks of trials. In the unbiased reward block, each correct response was

awarded 2 points. In the two biased reward blocks, a correct response was awarded 3 points for one category and 1 point for the other. Specifically, in one biased block, category A was awarded 3 points, while in another biased reward block, category B was awarded 3 points.

We implemented the CASANDRE model¹³ to compute metacognitive abilities from participants' categorisation and confidence responses. From the model fits, we extracted Signal-Detection-Theory-like parameters, which provided estimates of first-order decisions—sensitivity (i.e., d') and decision criterion (i.e., c)^{9,11}. Here, we expected decision criterion to shift toward the more likely or rewarded category (**Fig. 1c, Task 1**) and confidence to increase as orientations deviated from points of maximum overlap between categories (**Fig. 1d, Task 1**).

In Task 1, if prior or reward information were balanced, observers would achieve optimal performance by keeping the decision criterion at the intersection between the two categories, regardless of the stimulus uncertainty^{20,22}. Therefore, we used an embedded category task (Task 2) in Experiment 3 to assess how participants performed second-order perceptual decisions when first-order decisions could take into account sensory uncertainty alone. In Task 2, orientations were drawn from two embedded Gaussian distributions with means $m_A = m_B = 0^\circ$, and standard deviations $s_A = 3^\circ$ (Category A) and $s_B = 12^\circ$ (Category B) (**Fig. 1b, Task 2**). In this task, we expected the decision boundaries to shift outwards as the sensory uncertainty increased (**Fig. 1c, Task 2**). Hence, whereas in Task 1, decision shifts are driven by prior expectations or reward, in Task 2, decisions are driven by sensory uncertainty. Furthermore, we expected high confidence when category proportion favoured one category, and low confidence when the category proportions were similar (**Fig. 1d, Task 2**).

After applying exclusion criteria (see **Methods**, **Outlier removal**, and **Table 1**), the final sample included 42 non-autistic and 30 autistic participants in Experiment 1, 42 non-autistic and 27 autistic participants in Experiment 2, and 40 non-autistic and 26 participants in Experiment 3. A small number of additional participants were excluded from specific analyses (e.g., raw data, criterion), as detailed in the **Methods**.

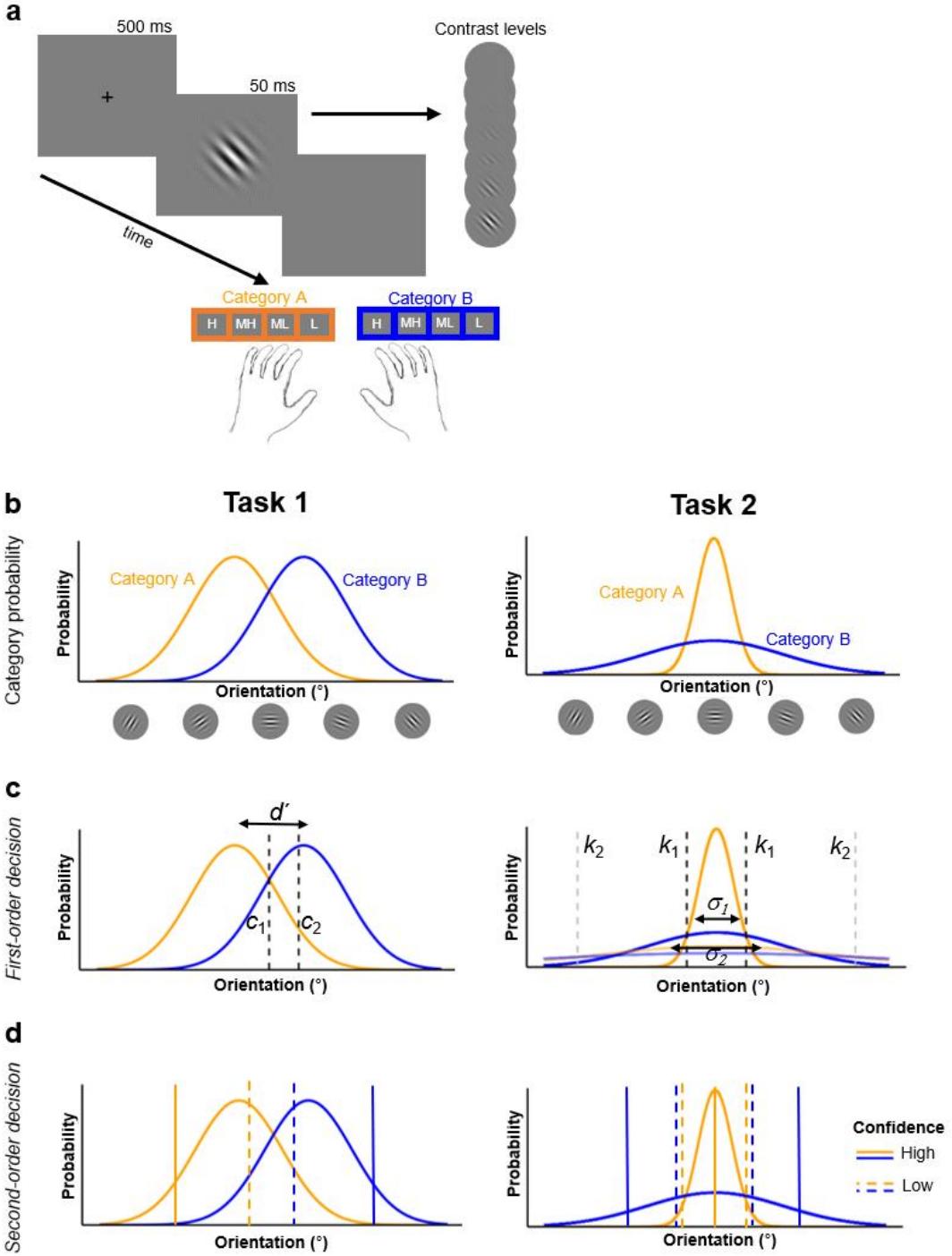


Fig. 1. Task descriptions. (a) Sequence of events within a trial in Experiments 1, 2, and 3, respectively, manipulating base rate, reward, and sensory uncertainty. For each trial, participants simultaneously reported the Gabor's category (Category A or Category B) based on its orientation, and their confidence level (high, medium-high, medium-low, low), using one of the eight keys, ranging from Category A highly confident to Category B highly confident. Stimulus contrast randomly varied between trials from a set of fixed values—0.004, 0.016, 0.033, 0.093, 0.18, 0.36, and 0.72. (b) Stimulus orientation distributions for each category in Task 1 (Experiments 1 and 2) and Task 2 (Experiment 3). (c) Internal representation of the category distributions for each task. In Task 1, the sensitivity d' represents the ability to separate the two categories, and the criterion c represents the adjustment of the decision criterion, from equal prior/reward (c_1) to prior/reward that favours Category A (c_2). In Task 2, the distributions with vivid colours represent the internal representations of the categories when the sensory noise is low, and the faded colours when the sensory noise is high. σ represents the standard deviation of the internal representation of Category A (i.e., combination of internal—inverse of d' —and external noises), leading to a narrow distribution when sensory noise is low (σ_1), and a wider when sensory noise is high (σ_2). k_1 represent the decision boundaries, shifting outwards when sensory noise is increasing (k_2). (d) Second-order decisions

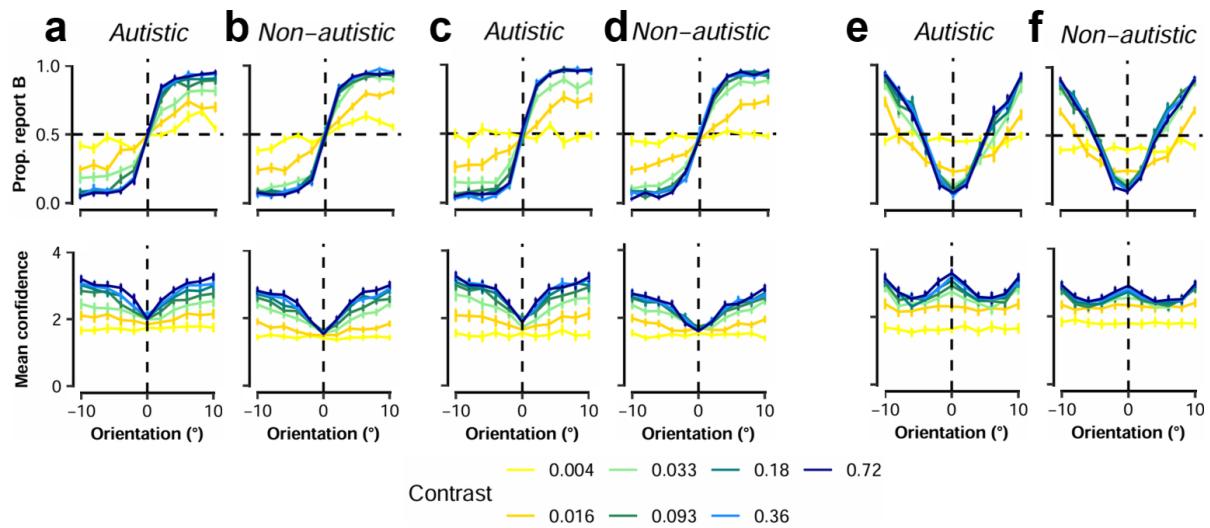
for each task. The vertical lines represent the confidence criterion c_c , indicating the stimulus information (i.e., orientation) required to report a specific level of confidence. The dashed lines represent the c_c for low confidence, and the solid line for high confidence. In Task 1, orientations around the category overlap (0°) are reported with low confidence for both categories, and orientations that deviate enough from the mean are reported with high confidence. In Task 2, orientations around the means of the two categories (0°) are associated with high confidence for Category A. Deviation from both sides of the category means gives similar evidence for both categories, leading to low confidence for A and B. Extreme deviation of orientation leads to report B with high Confidence.

1. Confidence ratings reflect sensory uncertainty for both groups

First, we examined how decision and confidence varied with stimulus orientation (11 orientations) and strength (7 contrast levels). **Fig. 2** illustrates the proportion of reporting Category B (top row) and the mean of confidence report (bottom row), as a function of orientation, contrast level, and group. Values reported are across base rate and reward blocks for Task 1 (**Fig. 2a-d**). For Experiments 1 and 2, manipulating category base rate and reward (**Fig. 2a-d, top row**), category report was characterised by a sigmoid shape, with a proportion of reports for Category B increasing as the orientation became more clockwise. The sigmoid was steeper with high contrasts, reflecting that category report was more sensitive to orientation as contrast increased. In Experiment 3, which manipulated sensory uncertainty, we observed that the probability of reporting Category B (wider distribution) increased as orientations deviated from 0° (i.e., mean of the narrow distribution), and the categorisation became more sensitive to orientation as contrast increased (**Fig. 2e-f, top row**, see **Supplementary Results** and **Supplementary Table 1**, for the statistical analyses). Importantly, Fazioli et al., (2023, 2025)^{18,19} and Fazioli et al., (in review) conducted optimal observer analyses on the data and showed similar first-order decisions for autistic and non-autistic groups when comparing them to optimal observers.

The variability in category reports is reflected in the confidence choices. **Fig. 2a-f, bottom row** illustrates how confidence increased as the probability of selecting a given category rose, with this relationship becoming sharper when stimulus strength increases (i.e., higher contrast). This consistent relationship between category and confidence choices indicates that participants can assess the reliability of their decision. To quantify how confidence choice was associated with stimulus information, we performed linear and quadratic mixed-effect models with binned orientation and contrast as within-subject factors and group as a between-subject factor, on confidence ratings.

Task 1



Task 2

Fig. 2. Category and confidence report data for (a, b) Experiment 1, prior manipulation, (c, d) Experiment 2, reward manipulation, and (e, f) Experiment 3, sensory uncertainty manipulation. The top row represents the proportion of reporting Category B (y-axis) as a function of orientation (x-axis) and contrast (line colour). The bottom row illustrates the Mean of confidence (y-axis) as a function of orientation (x-axis) and contrast level (line colour). In (a-d), data show means across participants and base rate/reward blocks. In (e, f), data show means across participants. Error bars represent \pm SE. The sample size consisted of 30 autistic and 41 non-autistic participants in (a, b), 27 autistic and 42 non-autistic participants in (c, d-top), 27 autistic and 41 non-autistic participants in (c, d-bottom), 26 autistic and 39 non-autistic participants in (e, f-top), and 26 autistic and 40 non-autistic participants in (e, f-bottom).

Confidence depends on the stimulus value and strength in both groups

The mixed-effect models investigating the effects of stimulus information (i.e., orientation, contrast) and group on confidence report included both linear and quadratic factors of orientation. Here, we focused on the overall V-shaped pattern and did not interpret linear terms separately.

Prior manipulation (Experiment 1)

In Experiment 1, as expected, confidence ratings were higher as contrast increased ($t(29.12) = 5.70, p < .001$), and this effect was more pronounced when orientations deviated from 0° , as indicated by the interaction between squared orientation and contrast ($t(59960) = 6.91, p < .001$). Participants reported overall higher confidence when base rate was unbalanced, compared to balanced ($t(59960) = 2.82, p = .005$) (see **Supplementary Figure 2a**), but overall confidence ratings did not significantly vary across groups ($t(47.70) = -1.54, p = .130$). All other main effects and interactions were not significant (see **Supplementary Results** and **Supplementary Table 2**).

Reward manipulation (Experiment 2)

Similarly to Experiment 1, confidence ratings were higher as contrast increased ($t(3.99) = 4.75, p = .009$), and this effect was more pronounced when orientations deviated from 0° , as indicated by the interaction between squared orientation and contrast ($t(51110) = 5.52, p < .001$). Participants were more confident in the unbalanced reward block than the balanced reward block ($t(54860) = 2.96, p = .003$) (see **Supplementary Figure 2b**). A significant three-way interaction between contrast, reward, and squared orientation ($t(55200) = 2.53, p = .012$) revealed that confidence ratings were more sensitive to stimulus value and strength when reward was unbalanced. However, these effects did not vary across groups, as all remaining effects, including the main effect of group ($t(25.54) = -0.71, p = .763$) and interactions, were not significant (see **Supplementary Results** and **Supplementary Table 2**).

Sensory uncertainty manipulation (Experiment 3)

As in the categorisation task in Experiments 1 and 2, confidence ratings in Task 2 increased with contrast ($t(5.01) = 5.65, p = .002$). Importantly, although overall confidence ratings did not significantly vary across groups ($t(15.42) = -0.10, p = .923$), autistic participants' confidence was more sensitive to stimulus value and strength, as indicated by a significant three-way interaction between group, squared orientation, and contrast ($t(98,770) = 2.99, p = .003$). All remaining effects and interactions were not significant (see **Supplementary Results** and **Supplementary Table 2**).

These results indicate that in both groups in Task 1 (Experiment 1 and 2), confidence ratings were driven by stimulus strength (i.e., contrast), an effect that depended on stimulus value (i.e., orientation). Importantly, the relation between stimulus value and strength varied between autistic and non-autistic individuals when first-order decision boundaries were adjusted for sensory uncertainty alone, as in Task 2 (Experiment 3). However, such relations rely on average adjustment of confidence based on participants' assessment of the strength of their first-order decision, but these relations do not reflect trial-to-trial variability in this assessment. Such variability, or meta-uncertainty, constitutes metacognitive abilities. To estimate each participant's meta-uncertainty, we next fitted the CASANDRE models of confidence¹³ that quantify their estimation of their own decision reliability, separately for each experiment.

2. Meta-uncertainty explains confidence reports in autism

According to the CASANDRE model, in perceptual decision-making, metacognitive ability in confidence report is determined by the observer's reliability of their uncertainty estimate on their perceptual choice. This well-established model explains previous works using both Tasks A and B^{13,20}. The model has two-stage processes (a first-order decision and a second-order confidence), and it is well rooted in traditional signal detection theory. The model separates discrimination abilities and response bias from meta-cognitive abilities. By fitting the model to the individual data, we estimated for each participant the following parameters: sensitivity (s), decision criterion (c_d), guess rate (g), confidence criterion (c_c), and meta-uncertainty (σ_m). Hence, σ_m provides a measure of estimation of internal noise that is independent of sensitivity and first-order decision. These parameters were optimised to minimise the negative log-likelihood and best capture participants' behavioural data (see **Methods, Computation model**).

2.1. Fit of meta-uncertainty model on category and confidence reports

See **Supplementary Methods** for details about comparisons between different variants of the meta-uncertainty model, and between models with meta-uncertainty as a free parameter (meta-uncertainty mode) and identical models with meta-uncertainty fixed at 0 (restricted model). Results demonstrate that meta-uncertainty plays a role in confidence reports in all experiments (see **Supplementary Results**).

The association of category and confidence reports with stimulus information was closely captured by the meta-uncertainty model in all experiments (see **Supplementary Figure 3**). To illustrate how the association between choice consistency and confidence was predicted by the meta-uncertainty model, **Fig. 3** displays the mean confidence as a function of the proportion of reporting B and contrast level, with observed (points) and predicted (solid lines) data for individual subjects, for the prior (**Fig. 3a-b**), reward (**Fig. 3c-d**), and sensory uncertainty (**Fig. 3e-f**) experiments. For the prior and reward experiments, we observed a single association ('U' shape) between confidence and category report across contrast conditions. This behaviour—captured by the meta-uncertainty model, as the predicted behaviour closely fitted the data—demonstrates participants' ability to assess the reliability of their decision. In the sensory uncertainty experiment, the more complex pattern of association between confidence and choice consistency was also captured by the meta-uncertainty model. Here, we noticed a greater variance in this association for autistic participants, as illustrated in **Fig. 3e-f**, suggesting more nuanced mapping of confidence on choice consistency.

Task 1

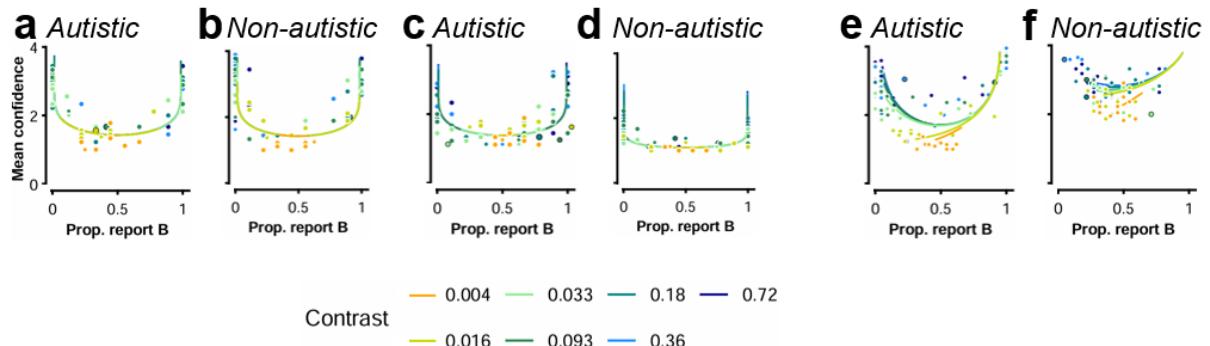


Fig. 3. Model fitting to category and confidence reports of a sampled participant from each group in each experiment. Mean of confidence (y-axis) as a function of proportion of reporting Category B (x-axis), contrast (colour), and group, for the prior (a, b), reward (c, d), and sensory uncertainty (e, f) experiments. Each subplot displays the observed and predicted behaviour for an individual participant. Data points illustrate choice behaviour, with size proportional to the number of trials. The solid lines represent the fit of the meta-uncertainty model using the maximum likelihood estimation method.

2.2. Comparable first-order sensitivity

We next analysed model-derived parameters reflecting first-order processes (i.e., sensitivity and decision criterion) in order to confirm that any differences in perceptual decisions between groups did not stem from atypical first-order decisions in autistic participants. **Fig. 4** illustrates perceptual sensitivity (top row) and decision criterion (bottom row) as a function of contrast level and group. In **Fig. 4a-b**, values are reported across base rate and reward blocks.

Prior manipulation (Experiment 1)

The ANOVA performed on sensitivity (s) revealed that sensitivity decreased with lower contrasts ($F(1.52, 106.28) = 105.47, p < .001, \eta_p^2 = .60$), indicating that the contrast variation was an effective manipulation of stimulus reliability (**Fig. 4a, top row**). None of the other effects were significant (see **Supplementary Results**), indicating that both groups exhibited a comparable perceptual sensitivity to the contrast levels.

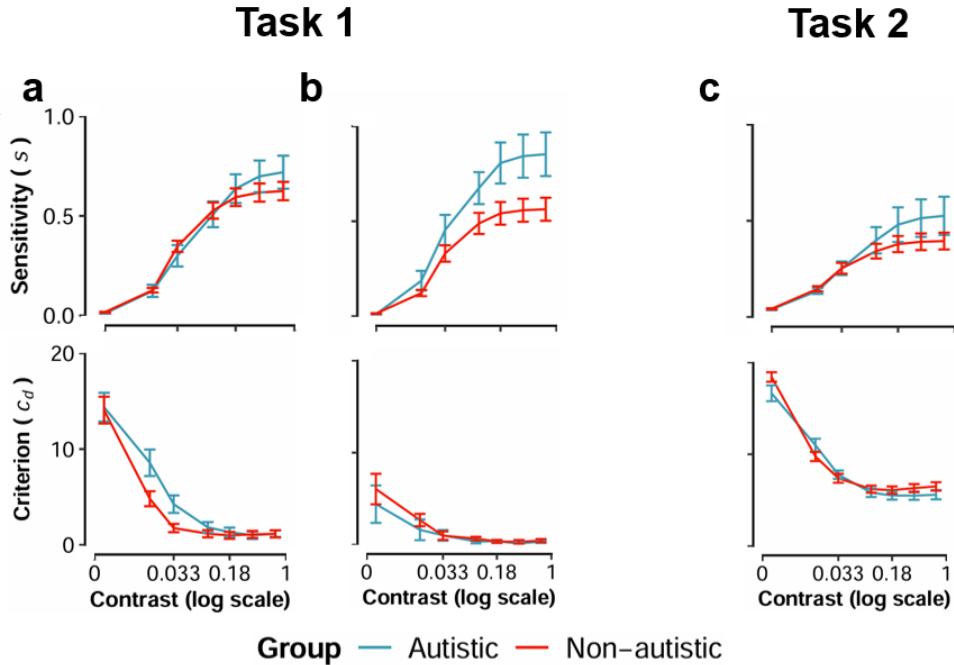


Fig. 4. Signal detection parameters for (a) Experiment 1 manipulating prior, (b) Experiment 2 manipulating reward, and (c) Experiment 3 manipulating sensory uncertainty. (a-c, top row) Perceptual sensitivity (s) as a function of contrast (x-axis) and group (line colour). The data are averaged across base rate in (a-top) and reward in (b-top). (a-c, bottom row) Decision criterion (c_d) as a function of contrast (x-axis) and group (line colour). The data displayed in (a, b, top row) are for the unbalanced base rate/reward blocks. All data points and bars show means across participants, and error bars represent \pm SE. The sample size consisted of 30 autistic and 42 non-autistic participants in (a), 27 autistic and 42 non-autistic participants in (b, top row), 27 autistic and 41 non-autistic participants in (b, bottom row), and 27 autistic and 40 non-autistic participants in (c).

Reward manipulation (Experiment 2)

The ANOVA performed on s revealed that sensitivity decreased as contrast decreased ($F(1.53, 102.80) = 72.51, p < .001, \eta_p^2 = .52$) (Fig. 34, top row). Furthermore, sensitivity was higher in the unbalanced, compared to the balanced reward blocks ($F(1, 67) = 8.83, p = .004, \eta_p^2 = .12$) (see **Supplementary Figure 4**), and this effect occurred at all contrast levels, except 0.033, as indicated by the significant interaction between reward and contrast ($F(1.31, 87.63) = 5.20, p = .017, \eta_p^2 = .07$). The two groups exhibited a comparable sensitivity ($F(1, 67) = 3.10, p = .083, \eta_p^2 = .04$) and all remaining effects were not significant (see **Supplementary Information**). Therefore, the two groups exhibited similar sensitivity to the contrast manipulation.

Sensory uncertainty manipulation (Experiment 3)

Sensitivity declined with decreasing contrast ($F(1.1, 65.94) = 64.69, p < .001, \eta_p^2 = .52$) (Fig. 4c, top row), and the two groups did not differ in sensitivity ($F(1, 60) = 0.94, p = .338, \eta_p^2 = .02$), across levels of contrast ($F(1.1, 65.94) = 2.28, p = .134, \eta_p^2 = .04$).

2.3. Comparable first-order decision criterion

Prior manipulation (Experiment 1)

The ANOVA performed on the decision criterion (c_d) revealed a greater criterion shift for unbalanced, compared to balanced base rate ($F(1, 70) = 124.56, p < .001, \eta_p^2 = .64$) (**Fig. 4a, bottom row**). Furthermore, criterion shift increased with decreasing contrast ($F(1.97, 138.18) = 116.12, p < .001, \eta_p^2 = .62$), and this only when base rate was unbalanced, as indicated by the interaction between base rate and contrast ($F(1.97, 138.18) = 116.12, p < .001, \eta_p^2 = .62$). Therefore, participants shifted their decision criterion toward the category with higher base rate probability as contrast decreased, and this in a comparable manner between groups, as the main effect of group was not significant ($F(1, 70) = 2.11, p = .151, \eta_p^2 = .03$), as well as all other effects (see **Supplementary Information**). Autistic individuals performed the categorisation task similarly to non-autistics, by exhibiting similar sensitivity and integration of prior information during first-order decisions.

Reward manipulation (Experiment 2)

Similarly, participants shifted their decision criterion more in the unbalanced, compared to balanced, reward block ($F(1, 69) = 19.61, p < .001, \eta_p^2 = .22$) (**Fig. 4b. bottom row**). Furthermore, criterion shift increased as contrast decreased ($F(1.43, 98.42) = 11.83, p < .001, \eta_p^2 = .15$), and this in a greater manner when reward was unbalanced ($F(1.43, 98.43) = 11.83, p < .001, \eta_p^2 = .15$), with no difference between groups ($F(1, 69) = 0.11, p = .738, \eta_p^2 < .001$). All other effects were not significant (see **Supplementary Information**).

Sensory uncertainty manipulation (Experiment 3)

The ANOVA performed on c_d revealed that c_d increased as contrast decreased ($F(2.47, 148.25) = 294.02, p < .001, \eta_p^2 = .83$) (**Fig. 4c, bottom row**). The two groups did not differ in the overall shift of decision criterion ($F(1, 60) = 0.53, p = .471, \eta_p^2 < .01$), however non-autistic participants tended to exhibit greater criterion shift in the contrast 0.004, as indicated by the significant interaction between group and contrast ($F(2.47, 148.25) = 3.71, p = .019, \eta_p^2 = .06$). This difference did not remain significant after correcting for multiple comparisons ($t(60) = 1.88, p = .065$).

Overall, these results confirm that in each experiment, the first-order decision reflects the specific contribution of one Bayesian component—prior, reward, or sensory uncertainty—to perceptual inference. Moreover, autistic individuals performed the first-order task similarly to

non-autistics. They exhibited comparable sensitivity to the orientation distributions and integrated all Bayesian components to the same extent.

3. Comparable confidence criterion across groups

In each experiment, the model estimated three confidence criteria, c_c , per participant, representing the amount of internal evidence a participant requires to report increasing levels of confidence. Specifically, c_c -low marks the boundary between the low and medium-low confidence keys, c_c -medium between medium-low and medium-high, and c_c -high between medium-high and high. A higher c_c value indicates a more conservative threshold, meaning the participant requires stronger evidence to shift the confidence level. **Fig. 5** illustrates the change in confidence criterion as a function of confidence level and group. In **Fig. 5a-b**, values are reported across base rate and reward blocks.

Prior manipulation (Experiment 1)

The ANOVA performed on c_c revealed that confidence criterion increased as confidence level increased ($F(1.52, 106.47) = 91.68, p < .001, \eta_p^2 = .57$)—where c_c -high was significantly higher than c_c -medium ($t(71) = 10.1, p < .001$), and c_c -medium was higher than c_c -low ($t(71) = 4.59, p < .001$) (**Fig. 5a**)—indicating that participants adopted more conservative thresholds when reporting higher confidence levels, demonstrating an appropriate use of the confidence rating scale. All remaining effects, including the effects of group ($F(1, 70) = 3.55, p = .064, \eta_p^2 = .05$) and base rate ($F(1, 70) = 0.04, p = .850, \eta_p^2 < .01$), were not significant (see **Supplementary Results**).

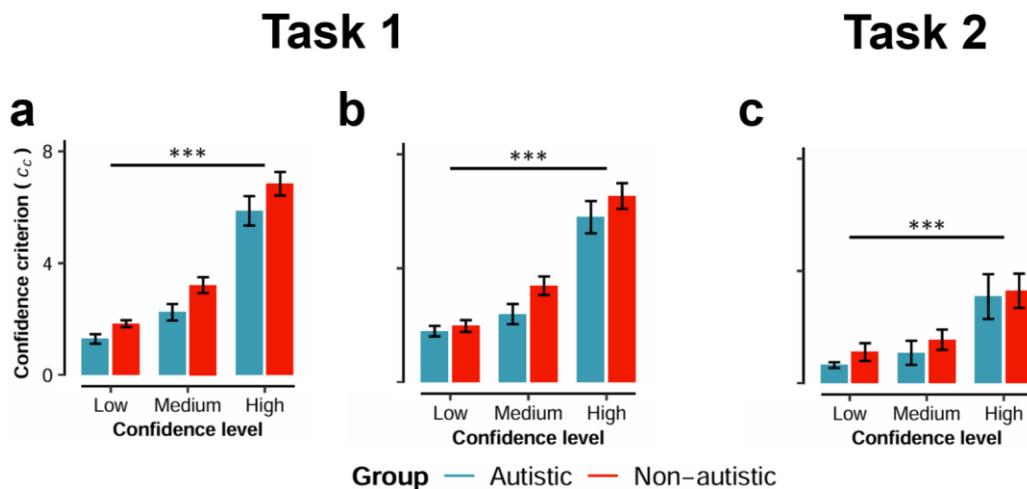


Fig. 5. Confidence criterion for (a) Experiment 1 manipulating prior, (b) Experiment 2 manipulating reward, and (c) Experiment 3 manipulating sensory uncertainty. Confidence criterion (c_c) as a function of

confidence level (x-axis) and group (bar colour). Confidence levels reflect the position of the criterion between two adjacent confidence keys: low/medium-low, medium-low/medium-high, and medium-high/high, respectively. The data are averaged across base rates in **(a)** and reward in **(b)**. Data points and bars show means across participants, and error bars represent \pm SE. The asterisks represent the main effect of confidence level evaluated with ANOVAs, $^{***}p < .001$. The sample size consisted of 30 autistic and 42 non-autistic participants in **(a)**, 27 autistic and 42 non-autistic participants in **(b)**, and 27 autistic and 40 non-autistic participants in **(c)**.

Reward manipulation (Experiment 2)

The ANOVA performed on the c_c showed that confidence criterion was more conservative as confidence level increased ($F(1.56, 104.32) = 53.15, p < .001, \eta_p^2 = .44$) (**Fig. 5b**), with a greater confidence criterion for c_c -high compared to c_c -medium ($t(68) = 8.00, p < .001$), and c_c -medium compared to c_c -low ($t(68) = 3.39, p = .004$). Importantly, confidence criterion did not vary between groups ($F(1, 67) = 1.62, p = .208, \eta_p^2 = .02$), and reward ($F(1, 67) = 3.14, p = .081, \eta_p^2 = .05$). All other effects were not significant (see **Supplementary Information**).

Sensory uncertainty manipulation (Experiment 3)

The ANOVA performed on c_c showed that confidence criterion increased with confidence level ($F(1.22, 73.43) = 20.17, p < .001, \eta_p^2 = .25$) (**Fig. 5c**), with a greater confidence criterion for c_c -high compared to c_c -medium ($t(61) = 4.73, p < .001$) and c_c -low ($t(61) = 4.73, p < .001$), while c_c -low and c_c -medium did not differ ($t(61) = 2.22, p = .091$). The confidence criterion did not vary between groups ($F(1, 60) = 0.50, p = .482, \eta_p^2 < .01$), or between confidence levels and groups ($F(1.22, 73.43) = 0.08, p = .833, \eta_p^2 < .01$).

Therefore, both groups similarly adjusted their confidence criterion by adopting a more conservative c_c when reporting higher confidence in their categorisation in all experiments, demonstrating a similar use of the confidence rating scale.

4. Meta-uncertainty in autism depends on first-order Bayesian source of uncertainty

Finally, we analysed the model estimation of meta-uncertainty (σ_m)—referring to the variability (uncertainty) in estimating the internal noise of the first-order decision variable—the estimate of perceptual metacognitive abilities (see **Methods**, **Computation model**, and **Data analyses**). A high σ_m value is associated with higher meta-uncertainty, and hence, lower metacognitive ability. To investigate whether and how cognitive abilities differed between groups, and the type of Bayesian information integrated during first-order decision, we conducted an ANOVA across experiments, with Bayesian information (prior, reward,

sensory uncertainty) and group as factors, on the σ_m . **Fig. 6** illustrates meta-uncertainty as a function of experiment and group, and means are reported across base rate and reward blocks. To investigate whether prior or reward conditions modulated meta-uncertainty, we also performed an ANOVA for each experiment with base rate/reward condition (balanced, unbalanced) and group as factors on the σ_m (see **Supplementary Results** and **Supplementary Figure 7**).

The two groups did not differ in overall meta-uncertainty ($F(1, 197) = 0.21, p = .648, \eta_p^2 < .01$), but rather during specific experiments, as indicated by the significant interaction between experiment and group ($F(2, 197) = 5.11, p = .007, \eta_p^2 = .05$) (**Fig. 6**). In the sensory uncertainty experiment, the autistic group exhibited lower meta-uncertainty compared to the non-autistic group ($t(78.8) = 2.39, p = .020, d = 0.59$), while in the prior experiment, the autistic group tended to exhibit higher meta-uncertainty ($t(32.3) = 1.92, p = .064, d = 0.49$). The ANOVA performed on meta-uncertainty from the prior experiment alone, investigating the difference between groups and base rate blocks, supported this tendency, showing that the autistic group exhibited greater meta-uncertainty compared to the non-autistic group ($F(1, 70) = 4.93, p = .030, \eta_p^2 = .07$) (see **Supplementary Results** and **Supplementary Figure 7**). The two groups did not differ in meta-uncertainty in the reward experiment ($t(58.8) = 0.35, p = .729$). Importantly, meta-uncertainty in the non-autistic group did not vary between experiments (post-hoc comparisons showed all $p > .05$). In contrast, in the autistic group, meta-uncertainty in the sensory uncertainty experiment was lower compared to the prior experiment ($t(197) = 4.03, p < .001$) and tended to be lower compared to the reward experiment ($t(197) = 2.24, p = .068$). The difference between the prior and reward experiments was not significant ($t(197) = 1.78, p = .179$). Overall, the main effect of experiment was significant, ($F(2, 197) = 4.57, p = .012, \eta_p^2 = .04$, but differences in meta-uncertainty between experiments were not significant after correcting for multiple comparisons: sensory uncertainty vs. prior, $t(130) = 2.32, p = .065$, sensory uncertainty vs. reward, $t(125) = 1.75, p = .164$, prior vs. reward, $t(129) = 0.86, p = .389$.

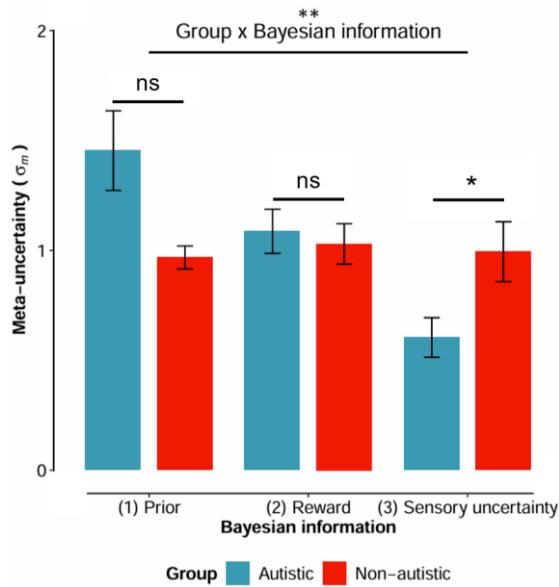


Fig. 6. Meta-uncertainty results per experiment. The meta-uncertainty σ_m (y-axis) as a function of experiment (x-axis) and group (bar colour). The data is averaged across base rate/reward blocks in Experiments 1 and 2. Bars show means across participants, and error bars represent \pm SE. The interaction between group and Bayesian component was evaluated using a between-subject ANOVA. The effects of group per experiment were evaluated using unpaired t-tests, with p-values corrected for multiple comparisons. ns: $p \geq .05$, $^{*}.05 > p \geq .01$, $^{**}.01 > p \geq .001$. The sample size consisted of 30 autistic and 42 non-autistic participants in Experiment 1, 27 autistic and 42 non-autistic participants in Experiment 2, and 27 autistic and 40 non-autistic participants in Experiment 3.

Because many participants participated in Experiment 3 after completing Experiment 1 and/or 2, we aimed to control for a possible training effect in Experiment 3. Therefore, to directly investigate if participation in previous experiments could improve metacognitive abilities, we tested whether meta-uncertainty varied between and within groups as a function of familiarity with the confidence report (i.e., whether participants performed Experiment 1 or 2 before completing Experiment 3). Results showed that in both groups, meta-uncertainty was the same between experienced and unexperienced participants (see **Supplementary Results** and **Supplementary Figure 5**), suggesting that group differences in meta-uncertainty cannot be explained by experimental training.

We conducted additional analyses to examine within-subject variability in meta-uncertainty between the prior and sensory-uncertainty experiments. The pattern of results indicates that within participants, non-autistics exhibit comparable meta-uncertainty between the two experiments, whereas autistics exhibit lower meta-uncertainty in the sensory uncertainty, compared to the prior experiment. These results support our general findings, demonstrating that meta-uncertainty in autistic participants depends on the Bayesian information integrated into the first-order decision (see **Supplementary Results** and **Supplementary Figure 6**).

Discussion

Alterations in self-monitoring and evaluation have been proposed to play a key role in several neurodevelopmental conditions^{2,3}, including autism²⁴. However, it remains unclear whether atypical self-monitoring in autism arises from differences in metacognitive ability, confidence bias, or is confounded by differences in first-order decisions. Using a computational modelling approach, we independently quantified these processes and identified a fundamental divergence in the factors that determine metacognitive ability in autistic versus non-autistic individuals.

Both autistic and non-autistic participants associated confidence with sensory uncertainty and adjusted their confidence criteria similarly. However, a key group difference emerged in *meta-uncertainty*—the computational estimate of uncertainty about internal noise. While meta-uncertainty remained stable in non-autistic participants, it varied in autistic participants depending on task manipulations of Bayesian components: metacognitive ability was enhanced (i.e., lower meta-uncertainty) when decisions relied solely on sensory evidence, but was reduced when prior information influenced the decision. Notably, group differences were specific to confidence reports. Perceptual sensitivity and decision criteria were comparable across groups, and both groups demonstrated similar integration of Bayesian information during first-order perceptual decision-making, as confirmed by comparisons with an ideal observer model in previous studies^{18,19} (Fazioli et al., *in review*). Together with the computational modelling results, these findings suggest atypical metacognitive abilities, even when sensory processing and first-order decision-making are indistinguishable.

In non-autistic individuals, metacognitive ability is typically considered a domain-general capacity—stable across tasks and sensory modalities^{25–28}. This consistency is also reflected in meta-uncertainty, which varies between individuals but is strongly correlated within individuals across sessions¹³. Our findings replicate this pattern: the average meta-uncertainty in the non-autistic group remained consistent across tasks. By contrast, metacognitive monitoring abilities in autism are context dependent. Rather than exhibiting a global reduction, autistic participants appear to monitor their own decisions more accurately when those decisions depend directly on sensory evidence and less accurately when prior knowledge or reward information influences the decision process. This suggests a reduced weighting of non-sensory information in self-evaluation—an effect that corresponds with claims of attenuated contextual integration in autistic perception^{17,29}. However, the present study, together with Fazioli et al.¹⁸, reveals that the attenuated effect of priors and context emerges at the metacognitive level, rather than at the level of first-order perceptual decisions¹⁷.

Metacognition and decision-making are dynamically coupled: confidence shapes learning, exploration, and behavioural adaptation^{30,31} and impacts subsequent choices and behaviour^{32,33}. Thus, differences in metacognitive ability may partly explain downstream behavioural differences. This is particularly relevant to sensory-related reactions in autism—an area that has received increasing attention in recent years—as atypical sensory reactivity, such as hypo- or hyper-responsiveness to sensory stimuli, has become recognized as a core feature of the autistic phenotype. While previous research has predominantly focused on sensory sensitivity and first-order decision-making^{16,17,34,35}, and some studies have proposed differences in higher-level expectations and priors^{29,36,37}, less attention was given to second-order perceptual processes, and the mechanisms underlying sensory reactivity in autism remain poorly understood. Our findings suggest that metacognitive monitoring capabilities may play an important and previously underappreciated role in shaping how autistic individuals engage with sensory input. Specifically, the reduced noise in estimating knowledge based on sensory evidence demonstrates a more accurate monitoring of sensory information in autistic individuals, suggesting a stronger bias towards sensory information at the metacognitive level, and reduced bias towards contextual information. These findings correspond to the claim of enhanced perception in autism³⁸. However, we suggest that enhanced abilities emerge at the second-order level rather than first-order sensitivity.

This bias towards sensory information may impact higher-order processes in decision-making, such as monitoring decisions that integrate priors—as found in Experiment 1—and therefore adjusting decision-making strategies based on prior information. These results suggest that the accumulated evidence for reduced prior updating in autism during first-order decisions^{39,40} may not originate in atypical first-order inference, but rather reflect weaker metacognitive calibration under prior-driven decisions. In particular, if autistic individuals fail to accurately evaluate the reliability of perceptual decisions when these are mainly guided by priors, contextual changes may not elicit the change in decision confidence that signals the need for adapting decision strategies through prior updating. Furthermore, a reduced ability to update priors based on environmental changes can impair the capacity to form predictive models of the environment, leading to an overestimation of volatility. This interpretation aligns with recent theoretical accounts proposing overestimation of volatility as a key feature of perceptual processing in autism^{36,37}.

Moreover, the present findings highlight the promise of this mechanistic framework to independently quantify first- and second-order perceptual processes—a novel contribution in developmental condition research. This computational approach of metacognitive ability

can capture individual differences overlooked by traditional accuracy-based measures, offering a framework to link self-monitoring with neural, developmental, and behavioural outcomes.

Finally, these results point to metacognition as a bridge for integrating perceptual processing and social accounts of autism. Metacognitive monitoring underpins one's capacity to interpret not only one's own mental states but also those of others; impairments in this domain may therefore contribute to challenges in communication and perspective-taking^{41–43}. Understanding how metacognition operates in autism may thus illuminate the developmental relationship between introspection, self-awareness, and theory of mind. This approach is relevant to the study of a broader range of mental disorders, as alterations in metacognitive computations are considered to play a critical role in many forms of psychopathology⁴⁴.

Methods

This study is based on data from a three-experiment project investigating perceptual decision-making in autism through categorisation of orientation tasks. The analyses on first-order decisions (i.e., category choice) were the object of Fazioli et al. (2023, 2025)^{18,19} and Fazioli et al. (in review). The present study employs the same experimental procedure and task design as those articles.

Participants

This study included 52 adults diagnosed with autism (41 males and 11 females) and 93 non-autistic individuals (18 males and 75 females). Participants chose between receiving monetary compensation (40 shekels/hour) or university credits (3 credits/hour). Autistic participants were recruited from a pool of participants regularly involved in research at the Department of Special Education. The two groups did not differ in age ($t(105) = .55, p = .59$), with a mean of $m = 26.70$ years old ($se = 0.86$) for the autistic group, and $m = 27.30$ ($se = 0.64$) for the non-autistic group. Intellectual Quotient (IQ) was assessed using the Test of Non-Verbal Intelligence (TONI-4), which measures cognitive functioning independent of language skills⁴⁵. The groups did not differ in IQ ($t(60.3) = .90, p = .37$), with a mean of $m = 99.3$ ($se = 11.40$) for the autistic group, and $m = 101.0$ ($se = 9.72$) for the non-autistic group. Autistic traits were measured using the Autistic Quotient (AQ) questionnaire, and a t-test ($t(64.9) = 6.97, p < .001$) revealed a significantly higher AQ score for the autistic group, $m = 27.0$ ($se = 8.11$), compared to the non-autistic group, $m = 16.7$ ($se = 6.69$). For each

participant, we maintained a minimum 24 hour-interval between experiments or experimental sessions.

The autism diagnosis was confirmed using standardised clinical assessments, including the DSM-5⁴, the Autism Diagnostic Interview (i.e., ADI-R52), and the Autism Diagnostic Observation Schedule (i.e., ASDOS-2). All participants completed the Community Assessment of Psychic Experiences (i.e., CAPE) and AQ questionnaires, in their preferred language (Hebrew or English), either during the clinical assessment or after the experimental sessions. Non-autistic individuals with a history of epilepsy or learning disorders were excluded from the study, as well as individuals diagnosed with autism who have known genetic disorders (e.g., Down syndrome).

The three experiments received ethical clearance from the Institutional Review Board at the University of Haifa under the reference number 046/20, and participants provided written informed consent before every experimental session.

Apparatus and Stimuli

Apparatus and stimuli. Participants were set in a dimly lit room in front of a computer. A chinrest was used to set viewing distance at 57 cm, and participants responded via a keyboard. See Fazioli et al. (2025) for information about the monitor and display background. Experimental design, tasks, and stimuli (**Fig. 1a**) were based on Qamar et al. (2013)²¹, Denison et al. (2017)²², and Adler and Ma (2018)²⁰. All stimuli were presented at the centre of the screen. Each trial began with a 500 ms fixation (black circle, 0.2° visual angle), followed by a 50 ms stimulus—a sinusoidal grating (Gabor patch) with a two-dimensional Gaussian envelope ($sd = 0.325^\circ$, 85% contrast, 3 cycles per degree). In each trial, the grating's orientation was randomly drawn from one of two Gaussian distributions, corresponding to the two stimulus categories (**Fig. 1a**). Observers were asked to report from which category they thought the stimulus belonged to, based on its orientation, and how confident they were about their answer. Following stimulus onset, they reported both their category choice (Category A or B) and their level of confidence using a 4-point scale using a single key. The confidence rating scale ranged from high-confidence Category A to high-confidence Category B (see **Fig. 1a**). To manipulate sensory uncertainty, we randomly varied stimulus contrast across trials, using seven fixed values (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72). The sensory uncertainty manipulation was used to 1) modulate the integration of prior and reward information into the perceptual decision in Experiments 1 and 2, and directly investigate the effect of sensory uncertainty on the decision boundaries in

Experiment 3, and 2) assess participants' ability to evaluate the reliability of their decision across different sensitivity levels— an estimate of metacognitive ability.

Categories. The stimulus categories were defined by continuous Gaussian orientation distributions. In Task 1 (Experiments 1 and 2), distributions were centred at $m_A = -4^\circ$ and $m_B = 4^\circ$ (relative to the horizontal line), with standard deviations of $sd_A = sd_B = 5^\circ$ (**Fig. 1b, Task 1**). In Task 2 (Experiment 3), we used embedded categories, a design allowing to test how changes in sensory uncertainty only influence perceptual decisions^{20,22,46}. Here, distributions had identical means, $m_A = m_B = 0^\circ$ (horizontal), but differing standard deviations, $sd_A = 3^\circ$ and $sd_B = 12^\circ$ (**Fig. 1b, Task 2**). These parameters were selected to yield an optimal accuracy rate of approximately 80%.

Blocks. Each experiment consisted of three blocks. In Experiment 1, we manipulated prior information by explicitly varying category base rate probabilities across blocks, with either balanced (B = 50% and A = 50%) or unbalanced (B = 25% and A = 75% or B = 75% and A = 25%) prior base rate between the two categories. In Experiment 2, we manipulated reward information by explicitly varying the number of points awarded for correct answers in each category across blocks, with balanced (B = 2 points and A = 2 points) or unbalanced (B = 1 point and A = 3 points or B = 3 points and A = 1 point) reward value between categories. In both experiments, the balanced block was always performed second. The order of the unbalanced blocks was counterbalanced between participants. In Experiment 3, as the sensory uncertainty was the main manipulation, there was no difference between the three experimental blocks.

Procedure and Design

Trainings. Each experiment started with extensive category (40 trials) and confidence (40 trials) training, with stimulus displayed for 300 ms at 100% contrast, see Fazioli et al. (2025) for more information.

Main experiment. Participants were explicitly introduced to the variation between each experimental block (e.g., base rate for Experiment 1 and reward points for Experiment 2) with a verbal explanation. At the beginning of each block, they were informed of the new base rate/reward condition, and performed a 40-trial practice session in which they reported both category and confidence. After each response, a text displayed the chosen category, along with a correctness auditory feedback. We ensured that participants reached around 70% accuracy during this practice session, reflecting that they were familiar enough with the

categories, the response keys, and the block conditions. Then, they completed the block of 280 test trials.

During the test blocks, no trial-to-trial feedback was given to prevent for feedback-based learning and ensure that decision boundaries were generated internally. However, participants received a summary of their categorisation accuracy every 50 trials to maintain engagement. In Experiments 2 and 3, participants also received information about the number of points earned in the previous 50 trials, and the total points accumulated throughout the experiment.

To ensure participants understood the main manipulations (i.e., base rate or reward), a “check question” was randomly introduced. In Experiment 1, participants were asked to gamble an amount (0-99 cents) on the chances for the next stimulus to belong to a specific category, with the remaining money assigned to the other category. They were informed that their prediction accuracy would influence a bonus added to their original compensation. In Experiment 2, participants were asked about how many points they would earn for correctly categorising a stimulus from a given category. Additionally, they were informed that the accumulated number of points earned during the experiment would determine a bonus added to their original compensation. No comprehension checks were needed in Experiment 3, so the same gambling question from Experiment 1 was used to maintain consistency. A reward system was also implemented to keep the same level of implication as in Experiments 1 and 2. Participants were informed that every correct answer was worth two points, and the total accumulated points would determine a bonus added to their compensation. In Experiments 1 and 2, participants completed 960 experimental trials over approximately 50 minutes. Preliminary data indicated that Experiment 3 was more susceptible to noise. Therefore, participants performed two separate sessions of 960 trials, with a minimum 24-hour gap between them.

Data analyses

We used MATLAB (R2024b) to fit the computational model to our data. Statistical analyses were conducted in R (4.4.1).

Based on previous analyses, we assumed a symmetry in participants’ criterion shift between opposite base rate/reward¹⁸ blocks. Therefore, before implementing the model, we converted the responses from blocks where Category A had low base rate/reward, in order to combine the trials with the other unbalanced block for a single model fit. We reversed stimulus category and stimulus responses and multiplied the stimulus orientation by -1.

Outlier removal

In all Experiments, participants with an accuracy below 0.6 at the three highest contrast levels and across blocks, were excluded from all analyses. In Experiment 3, we also removed participants with extreme criterion shifts ($k > 100$) or a sensory uncertainty ($\sigma > 100$) from all analyses¹⁸. We also excluded participants who used one key per category, indicating that they reported categorisation choice only, and participants who exhibited a meta-uncertainty that fell above the third quartile plus three times the interquartile range from all analyses. Additionally, we excluded participants showing an overall negative decision criterion from the decision criterion analyses. Finally, participants who did not have trials in all combinations of binned orientation and contrast were automatically excluded from the behavioural data analyses (category and confidence report).

Computational model

To estimate meta-cognitive abilities, we fitted a recent computational model (the 'CASANDRE' or 'confidence as a noisy decision reliability estimate' model)¹³ that was shown to explain well behavioural confidence reports in previous studies using the same basic stimuli and task¹³. The model estimates meta-uncertainty, a metacognitive parameter reflecting how precisely an individual can assess their decision reliability. The model assumes that on each trial, an observer estimates the reliability of their decision (makes a confidence decision, V_c) by comparing the absolute distance between the decision variable V_d (i.e., strength of sensory evidence) and the contrast-specific decision criterion c_d , and normalising it by $\hat{\sigma}_d$, the estimate of the dispersion of V_d (Eq. 1). Here, V_d is derived from a normal distribution centred on the true stimulus value, with a variability given by sensory noise (i.e., inverse of sensitivity). The absolute difference between V_d and V_c reflects the strength of the decision, with a higher value indicating stronger evidence. Therefore, V_c can be explained as the strength of evidence for the choice, scaled by how noisy the internal system is perceived to be. Indeed, this framework assumes that observers don't have access to their actual sensory uncertainty and estimate for every decision. This estimate $\hat{\sigma}_d$ is modelled as a random variable drawn from a lognormal distribution with a mean of σ_d (i.e., the true sensory noise), and a trial-to-trial variability of σ_m . Finally, confidence rating was obtained by comparing V_d to a fixed confidence criterion c_c . Therefore, the noise when estimating decision reliability mainly comes from variability in assessing sensory uncertainty, also called meta-uncertainty σ_m . A larger σ_m indicates greater variability in estimating internal

noise, and therefore, lower metacognitive abilities. Additionally, a guess rate (g) parameter is included to account for random response.

$$V_c = \frac{|V_d - c_d|}{\hat{\sigma}_d} \quad (1)$$

For each participant in each contrast level, the model estimates sensitivity (s) and decision criterion (c_d). Additionally, for each participant across all contrast levels, the model estimates guess-rate (g), confidence criterion (c_c), and meta-uncertainty (σ_m).

To apply the model, we modelled sensitivity using a Naka-Rushton function, defined by a maximum sensitivity R_{max} , a semi-saturation constant C_{50} , and a slope n (Eq. 2). Therefore, the model contained 15 free parameters: GR, meta-uncertainty (σ_m), three sensory parameters (R_{max} , C_{50} , n), seven decision criteria (c_d), and three confidence criteria (c_c , one less than the number of confidence levels). Each parameter was optimised to minimise negative log-likelihood, using the MATLAB fmicon function. The fitting procedure was structured in three nested loops, and the best fit was selected based on the lowest log-likelihood.

$$s = \frac{R_{max} * C^n}{C^n + C_{50}^n} \quad (2)$$

The starting values of GR, σ_m , R_{max} , C_{50} , n , and c_c were respectively: 0.01, 0.5, 1.5, 0.3, 2, 1. We defined strict lower and upper bounds for each parameter to ensure valid estimates: $0 < GR < 0.1$; $0.1 < \sigma_m < 5$; $0.005 < R_{max} < 5$, $0.5 < n < 5$; $0.005 < C_{50} < 1$; $0 < c_c < 10$.

In Task 1, the starting values of the decision criteria c_d were set at 0. For the trials from unbalanced blocks (i.e., unequal reward or prior between categories), the boundaries were $-20 < c_d < 20$. For the trials from the balanced block (i.e., equal reward or prior between categories), the boundaries were $-0.002 < c_d < 0.002$, as the criterion was not expected to vary when sensory evidence decreased, in order to reduce the number of free parameters to 12.

In Task 2, the observer sets two decision boundaries to distinguish between the narrow category A and the broad category B (Fig. 1c, Task 2). To reduce the number of free parameters, we assumed these boundaries to be symmetrical around zero degrees. Therefore, we estimated the lower value of c_d for each contrast, and multiplied it by minus

one to estimate the upper value. The boundaries were set at $-20 < c_d < 0$, and the starting values were set from -5 (high contrast) to -11 (low contrast). Therefore, similar to the unbalanced trials of Task 1, we used 15 free parameters for modelling data from Task 2.

Model comparison

We started by fitting the original model to the data. In that version, sensitivity was estimated separately for each contrast level (i.e., 7 values), using a signal-detection-theory-like model. For Task 1, the model did not conditionally constrain c_d , and for Task 2, both boundaries for c_d were estimated. This resulted in a total of 19 free parameters for Task 1, and 26 for Task 2. To reduce the number of free parameters and minimise the risk of overfitting, we tested three alternative model variants, where estimates for sensitivity and decision criterion were reduced. The model described above, using the Naka-Rushton function to estimate sensitivity, demonstrated the best fit (see **Supplementary Results**). The results reported in the main text are based on the parameters extracted from this model. The description and comparison between models are provided in the **Supplementary Methods**, **Supplementary Results**, and **Supplementary Figure 1**.

Statistical analyses

Behavioural data

Categorisation task

For each experiment, we investigated how the category choice varied across contrast and orientations. For Experiments 1 and 2, we conducted $2 \times 7 \times 11 \times 2$ linear mixed-effect models with group (non-autistic, autistic) as a between-subject factor, and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72), orientation (-10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10), and base rate / reward (balanced, unbalanced) as within-subject factors, on the proportion of reporting Category B. For Experiment 3, we conducted a $2 \times 7 \times 11$ linear and quadratic mixed-effect model to account for the V-shaped pattern of response. The model included group (non-autistic, autistic) as a between-subject factor, contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72), and orientation (-10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10) as within-subject factors, as well as the squared orientation factor, and was performed on the proportion of reporting Category B (see **Supplementary Results**, **Supplementary Table 1**)

Confidence task

We investigated how the confidence report was influenced by the different manipulations for each experiment. In Experiments 1 and 2, we conducted $2 \times 7 \times 11 \times 2$ linear and quadratic mixed-effect models with group (non-autistic, autistic) as a between-subject factor, and

contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72), orientation (-10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10), and base rate/reward block (balanced, unbalanced) as within-subject factors, on the confidence report. A squared orientation factor was added in each model to account for the V-shaped behaviour. In Experiment 3, we conducted a $2 \times 7 \times 11$ linear mixed-effect model with group (non-autistic, autistic) as a between-subject factor, and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72), and orientation (-10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10) as within-subject factors, on the confidence report.

Sensitivity and decision criterion

We used the parameters from the Naka-Rushton fitting to estimate the sensitivity s_i for each level of contrast C_i for each participant (Eq. 3). In Experiments 1 and 2, we performed $2 \times 2 \times 7$ mixed-design ANOVAs with group (non-autistic, autistic) as a between-subject factor, and prior/reward (balanced, unbalanced) and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72) as within-subject factors on the s and c_d . In Experiment 3, we performed 2×7 mixed-design ANOVAs with group (non-autistic, autistic) as a between-subject factor, and contrast (0.004, 0.016, 0.033, 0.093, 0.18, 0.36, 0.72) as a within-subject factor on the s and c_d .

$$s_i = \frac{R_{max} * C_i^n}{C_i^n + C_{50}^n} \quad (3)$$

Confidence criterion

Because there were four confidence levels, there were three confidence-level boundaries: 1=between low and mid-low, 2 = between mid-low to mid-high, 3 = between mid-high and high). To investigate the shift in confidence criterion c_c , we performed a $2 \times 3 \times 2$ mixed-design ANOVA with group (non-autistic, autistic) as a between-subject factor and confidence-level boundaries (1, 2, and 3) and base-rate block (balanced, unbalanced) as within-subject factors on the c_c . In Experiment 2, we performed a similar $2 \times 3 \times 2$ mixed-design ANOVA with group, confidence-level boundary, and reward block (balanced, unbalanced). In Experiment 3, we performed a 2×3 mixed-design ANOVA with group (non-autistic, autistic) as a between-subject factor and confidence level (1, 2, 3) as a within-subject factor on the c_c .

Meta-uncertainty

To investigate whether meta-uncertainty differed between groups, and whether the type of Bayesian information involved in perceptual decisions affected the metacognitive abilities, we performed a 2×3 between-subjects ANOVA with group (non-autistic, autistic) and

experiment (prior, reward, sensory uncertainty) as between-subject factors, on the meta-uncertainty.

To control for potential improvement in metacognitive abilities over time among participants who completed multiple experiments, we directly tested whether meta-uncertainty in Experiment 3 varied as a function of familiarity with the task, by conducting a 2×2 between-subject ANOVA on the meta-uncertainty in Experiment 3, with group (non-autistic, autistic) and previous participation (with, without) as between-subject factors. See **Supplementary Results** and **Supplementary Figure 4**.

To investigate the difference between groups and block conditions (prior, reward) in metacognitive abilities for Experiments 1 and 2, we conducted for each experiment a 2×2 mixed-design ANOVA with group (non-autistic, autistic) as a between-subject factor, and prior/reward block (balanced, unbalanced) as a within-subject factor on the σ_m . The results are described in the **Supplementary Results** and **Supplementary Figure 6**.

Guess rate

In Experiments 1 and 2, we performed a 2×2 mixed-design ANOVAs with group (non-autistic, autistic) as a between-subject factor, and prior/reward block (balanced, unbalanced) as a within-subject factor on g . In Experiment 3, we conducted an unpaired t-test on g with group (non-autistic, autistic) as the between-subject factor. The results are displayed in the **Supplementary Information**.

Significant effects from the ANOVAs were further investigated using paired and unpaired t-tests as appropriate to elucidate the nature of the observed differences. Bonferroni corrections were applied to control for multiple comparisons. Effect sizes were calculated using partial eta square (η_p^2) for ANOVAs and Cohen's standardised mean difference (d) for t-tests.

Data availability

The datasets supporting the conclusions of this article are available in the Open Science Framework (OSF) repository: <https://osf.io/kp7ca/>

Code availability

The code supporting these findings will be available on the Open Science Framework (OSF) repository: <https://osf.io/kp7ca/>

Acknowledgments

We would like to thank all the participants who took part in this project, as well as Dr. Marissa Hartston for recruiting the autistic participants. We also would like to thank our funding sources: The Israel Science Foundation (Grant No. 1459/24 to A. Yashar, and 2025/25 to B-S. Hadad) and the U.S.-Israeli Binational Science Foundation (Grant No. 234/2020 to B-S. Hadad).

Competing interest statement

The authors declare no conflict of interest.

Authors contribution

AY, B-S.H and RD, conceptualized and designed the study. LF conducted and collected the data. LF and AY analyzed the data. The original draft was written by LF and AY and reviewed by all authors.

References

1. Flavell, J. H. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.* **34**, 906–911 (1979).
2. Seow, T. X. F., Rouault, M., Gillan, C. M. & Fleming, S. M. How Local and Global Metacognition Shape Mental Health. *Biol. Psychiatry* **90**, 436–446 (2021).
3. Wise, T., Robinson, O. J. & Gillan, C. M. Identifying Transdiagnostic Mechanisms in Mental Health Using Computational Factor Modeling. *Biol. Psychiatry* **93**, 690–703 (2023).
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. (American Psychiatric Association, Arlington, VA, 2022).
5. Carpenter, K. L. & Williams, D. M. A meta-analysis and critical review of metacognitive accuracy in autism. *Autism* **27**, 512–525 (2023).
6. Maras, K., Norris, J. E. & Brewer, N. Metacognitive Monitoring and Control of Eyewitness Memory Reports in Autism. *Autism Res.* **13**, 2017–2029 (2020).
7. DeBrabander, K. M., Pinkham, A. E., Ackerman, R. A., Jones, D. R. & Sasson, N. J. Cognitive and Social Cognitive Self-assessment in Autistic Adults. *J. Autism Dev. Disord.* **51**, 2354–2368 (2021).

8. Nelson, T. O. A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull.* **95**, 109–133 (1984).
9. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **9**.
10. Mamassian, P. Visual Confidence. *Annu. Rev. Vis. Sci.* **2**, 459–481 (2016).
11. Maniscalco, B. & Lau, H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**, 422–430 (2012).
12. Fleming, S. M. HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci. Conscious.* **2017**, (2017).
13. Boundy-Singer, Z. M., Ziembra, C. M. & Goris, R. L. T. Confidence reflects a noisy decision reliability estimate. *Nat. Hum. Behav.* **7**, 142–154 (2022).
14. Knill, D. C. & Richards, W. *Perception as Bayesian Inference*. (1996).
15. Mamassian, P., Landy, M. & Maloney, L. T. *Bayesian Modelling of Visual Perception*. (2002).
16. Lawson, R. P., Rees, G. & Friston, K. J. An aberrant precision account of autism. *Front. Hum. Neurosci.* **8**, (2014).
17. Pellicano, E. & Burr, D. When the world becomes ‘too real’: a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 504–510 (2012).
18. Fazioli, L., Hadad, B.-S., Denison, R. N. & Yashar, A. Suboptimal but intact integration of Bayesian components during perceptual decision-making in autism. *Mol. Autism* **16**, 2 (2025).
19. Fazioli, L., Hadad, B.-S., Denison, R. & Yashar, A. Intact Bayesian perceptual decision making and metacognition in autism. *J. Vis.* **23**, 5283 (2023).
20. Adler, W. T. & Ma, W. J. Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLOS Comput. Biol.* **14**, e1006572 (2018).
21. Qamar, A. T. *et al*. Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proc. Natl. Acad. Sci.* **110**, 20332–20337 (2013).
22. Denison, R. N., Adler, W. T., Carrasco, M. & Ma, W. J. Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proc. Natl. Acad. Sci.* **115**, 11090–11095 (2018).
23. Navajas, J., Bahrami, B. & Latham, P. E. Post-decisional accounts of biases in confidence. *Curr. Opin. Behav. Sci.* **11**, 55–60 (2016).
24. Bednarz, H. M., Trapani, J. A. & Kana, R. K. Metacognition and behavioral regulation predict distinct aspects of social functioning in autism spectrum disorder. *Child Neuropsychol.* **26**, 953–981 (2020).
25. Faivre, N., Filevich, E., Solovey, G., Kühn, S. & Blanke, O. Behavioural, modeling, and electrophysiological evidence for domain-generality in human metacognition. Preprint at <https://doi.org/10.1101/095950> (2016).
26. Morales, J., Lau, H. & Fleming, S. M. Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *J. Neurosci.* **38**, 3534–3546 (2018).

27. Carpenter, J. *et al.* Domain-general enhancements of metacognitive ability through adaptive training. *J. Exp. Psychol. Gen.* **148**, 51–64 (2019).

28. Rouault, M., McWilliams, A., Allen, M. G. & Fleming, S. M. Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personal. Neurosci.* **1**, e17 (2018).

29. Friston, K. J., Lawson, R. & Frith, C. D. On hyperpriors and hypopriors: comment on Pellicano and Burr. *Trends Cogn. Sci.* **17**, 1 (2013).

30. Lee, D. G. & Hare, T. A. Value certainty and choice confidence are multidimensional constructs that guide decision-making. *Cogn. Affect. Behav. Neurosci.* **23**, 503–521 (2023).

31. Yeung, N. & Summerfield, C. Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1310–1321 (2012).

32. Rahnev, D., Koizumi, A., McCurdy, L. Y., D’Esposito, M. & Lau, H. Confidence Leak in Perceptual Decision Making. *Psychol. Sci.* **26**, 1664–1680 (2015).

33. Purcell, B. A. & Kiani, R. Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proc. Natl. Acad. Sci.* **113**, (2016).

34. Hadad, B.-S. & Yashar, A. Sensory Perception in Autism: What Can We Learn? *Annu. Rev. Vis. Sci.* **8**, 239–264 (2022).

35. Mottron, L., Dawson, M., Soulières, I., Hubert, B. & Burack, J. Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *J. Autism Dev. Disord.* **36**, 27–43 (2006).

36. Lawson, R. P., Mathys, C. & Rees, G. Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* **20**, 1293–1299 (2017).

37. Van De Cruys, S. *et al.* Precise minds in uncertain worlds: Predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).

38. Mottron, L., Dawson, M., Soulières, I., Hubert, B. & Burack, J. Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *J. Autism Dev. Disord.* **36**, 27–43 (2006).

39. Sapey-Triomphe, L., Timmermans, L. & Wagemans, J. Priors Bias Perceptual Decisions in Autism, But Are Less Flexibly Adjusted to the Context. *Autism Res.* **14**, 1134–1146 (2021).

40. Twito, R., Hadad, B. & Szpiro, S. Is she still angry? Intact learning but no updating of facial expressions priors in autism. *Autism Res.* **17**, 934–946 (2024).

41. Grainger, C., Williams, D. M. & Lind, S. E. Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgement of confidence accuracy. *Conscious. Cogn.* **42**, 65–74 (2016).

42. Nicholson, T., Williams, D. M., Lind, S. E., Grainger, C. & Carruthers, P. Linking metacognition and mindreading: Evidence from autism and dual-task investigations. *J. Exp. Psychol. Gen.* **150**, 206–220 (2021).

43. Fleming, S. M. Metacognition and Confidence: A Review and Synthesis. *Annu. Rev. Psychol.* **75**, 241–268 (2024).

44. Rouault, M., Seow, T., Gillan, C. M. & Fleming, S. M. Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biol. Psychiatry* **84**, 443–451 (2018).

45. Goldberg Edelson, M., Edelson, S. M. & Jung, S. Assessing the Intelligence of Individuals with Autism: A Cross-Cultural Replication of the Usefulness of the TONI. *Focus Autism Dev. Disabil.* **13**, 221–227 (1998).

46. Qamar, A. T. *et al.* Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proc. Natl. Acad. Sci.* **110**, 20332–20337 (2013).

Table

	Overall n	Category report	Confidence report	Sensitivity	Decision criterion	Confidence criterion	Meta-uncertainty	Guess rate
Prior experiment	$n_{autistic} = 34$ $n_{non-autistic} = 49$	$n_{autistic} = 30$ $n_{non-autistic} = 41$	$n_{autistic} = 30$ $n_{non-autistic} = 41$	$n_{autistic} = 30$ $n_{non-autistic} = 42$				
Reward experiment	$n_{autistic} = 32$ $n_{non-autistic} = 48$	$n_{autistic} = 27$ $n_{non-autistic} = 42$	$n_{autistic} = 27$ $n_{non-autistic} = 41$	$n_{autistic} = 27$ $n_{non-autistic} = 42$	$n_{autistic} = 27$ $n_{non-autistic} = 41$	$n_{autistic} = 27$ $n_{non-autistic} = 42$	$n_{autistic} = 27$ $n_{non-autistic} = 42$	$n_{autistic} = 27$ $n_{non-autistic} = 42$
Sensory uncertainty experiment	$n_{autistic} = 34$ $n_{non-autistic} = 44$	$n_{autistic} = 26$ $n_{non-autistic} = 39$	$n_{autistic} = 26$ $n_{non-autistic} = 40$	$n_{autistic} = 27$ $n_{non-autistic} = 40$				

Table 1. Description of the sample size in the three experiments for the overall sample, and statistical analyses performed on each estimate: category report, confidence report, sensitivity, decision criterion, confidence criterion, meta-uncertainty, guess rate.

Supplementary Information

Enhanced metacognition in autism when perceptual decisions rely solely on sensory evidence

Laurina Fazioli¹, Bat-Sheva Hadad¹, Rachel N. Denison² and Amit Yashar¹

Supplementary methods

Model fitting

For each experiment, we performed four model fittings, varying the method for estimating sensitivity to reduce the number of free parameters, while preserving a good fit to the data.

Original model

In the original model, stimulus sensitivity s_i was estimated separately for each contrast (i.e., 7 levels), using a Signal Detection Theory (SDT)-based model of choice and confidence. Each s_i was treated as a free parameter, optimised via maximum likelihood to best predict the participant's decision variable V_d . To reduce the number of free parameters, we implemented alternative models in which sensitivity was fitted with parametric functions during the optimisation of V_d .

Linear model

In the linear model, changes in sensitivity across contrast were modelled with a linear function for each participant. Here, C_i represents stimulus contrast, a the slope, and b the intercept (**Eq. 1**). Two free parameters (a and b) were estimated per participant.

$$s_i = a * C_i + b \quad (1)$$

Gaussian model

Visualisation of sensitivity revealed that changes did not follow a strict linear progression. We used a Gaussian cumulative distribution function (CDF) to capture the nonlinear behavior (**Eq. 2**). In this equation, μ represents the inflection point (i.e., center of the curve), σ the spread (i.e., slope), and ϕ the cumulative normal distribution. Two free parameters were estimated (μ and σ) per participant.

$$s_i = \Phi\left(\frac{C_i - \mu}{\sigma}\right) \quad (2)$$

Naka-Rushton model

Finally, to account for saturation of sensitivity at high contrast levels, we used a Naka-Rushton function to model sensitivity (**Eq. 3**). Here, C_i is the stimulus contrast, R_{\max} the asymptotic maximum sensitivity, C_{50} the contrast at which sensitivity reaches half of R_{\max} , and n the slope parameter controlling the steepness of the function.

$$s_i = \frac{R_{\max} * C_i^n}{C_i^n + C_{50}^n} \quad (3)$$

Statistical analyses

Model comparison

We fitted all four models (i.e., original, linear, Gaussian, Naka-Rushton) to each experiment's datasets. Each fit produced one set of parameters for the meta-uncertainty model (meta-uncertainty as a free parameter), and another set for the reduced model (meta-uncertainty fixed at 0). For each model, the negative likelihood (NLL) was computed to quantify how the model predicted the observed data. From the NLL values, we calculated the corrected Akaike Information Criterion ($AICc$) to evaluate model quality while accounting for the number of free parameters (k) and trials (n) (**Eq. 4**). For each model, we performed t-tests on the $AICc$ to assess whether the model including the meta-uncertainty parameter outperformed the model without it.

$$AICc = (2k + 2NLL) + \frac{2k*(k+1)}{n-k-1} \quad (4)$$

To compare the different variants (i.e., linear, Gaussian, Naka-Rushton) in each experiment, we computed for each participant the difference between the $AICc$ of the original model and the $AICc$ of each variant. High values in the resulting $\Delta AICc$ indicate a better fit to the data. We then performed mixed-design ANOVAs on the $\Delta AICc$ to find the best-fitting model, with Model (linear, Gaussian, Naka-Rushton) as a within-subject factor, and group (autistic, non-autistic) as a between-subject factor. In Experiments 1 and 2, the factor base rate/reward block (balanced, unbalanced) was added to the ANOVAs. Finally, we performed one-sample t-tests on $\Delta AICc$ values corresponding to the model variant with the highest $\Delta AICc$. This

tested whether the mean $\Delta AICc$ was significantly greater than 0, indicating that the selected variant provided a better fit than the original model.

After selecting the variant model that outperforms the original model based on the $\Delta AICc$, we evaluated, for the selected variant, whether including the meta-uncertainty as a free parameter improved model fit. For each experiment, we calculated the $AICc$ gain, defined as the difference between the $AICc$ of the restricted model and the $AICc$ of the meta-uncertainty model, where a positive value indicates that including the meta-uncertainty as a free parameter improves model fit. We excluded participants with extreme $AICc$ gain values ($> 3 sd$): one autistic participant in Experiment 1, one non-autistic participant in Experiment 2, and three autistic and four non-autistic participants in Experiment 3, resulting in a final sample of 71, 68, and 55 participants, respectively. Then, we conducted one-sample t-tests against 0 to test whether models including the meta-uncertainty models systematically outperformed the restricted models, separately for each experiment.

Supplementary results

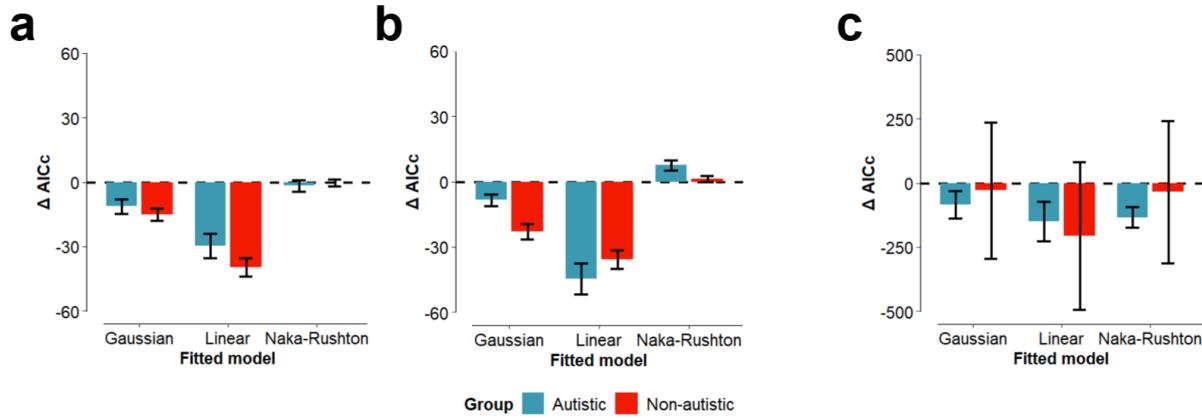
Model comparison

We first compared the model fits for the different variants of the contrast sensitivity function (i.e., linear, Gaussian, Naka-Rushton) across experiments. **Supplementary Figure 1** illustrates $\Delta AICc$ as a function of model fitted per groups. In Task 1 (plots a-b), values are reported across base rate and reward blocks.

Prior experiment

The ANOVA performed on the $\Delta AICc$ revealed a main effect of model $F(1.39, 97.38) = 35.93, p < .001, \eta_p^2 = .34$, with the Naka-Rushton model being a better predictor than both the Gaussian ($t(77) = 5.40, p < .001$) and the linear ($t(77) = 7.24, p < .001$) models (**Supplementary Figure 1**). Furthermore, the main effect of base rate was significant ($F(1, 70) = 4.99, p = .029, \eta_p^2 = .07$), indicating that the models predicted better the data from the balanced, compared to the unbalanced blocks. The interaction between model and base rate was significant ($F(1.96, 137.01) = 5.72, p = .004, \eta_p^2 = .08$), and caused by a main effect of base rate only in the linear model ($t(76) = 3.63, p < .001$), compared to the Gaussian ($t(76) = 1.34, p = .182$) and Naka-Rushton ($t(76) = 0.92, p = .359$) models. All other effects were not significant, including the main effect of group ($F(1, 70) = 1.08, p = .303, \eta_p^2 = .02$), the interaction between group and model ($F(1.39, 97.38) = 0.99, p = .348, \eta_p^2 = .01$), and the three-way interaction ($F(1.96, 137.01) = 2.53, p = .084, \eta_p^2 = .04$). The one-sample t-test performed on $\Delta AICc$ of the Naka-Rushton model revealed no significant difference from 0 ($t(71) = 0.49, p = .628$), that this model did not outperform the original one. Therefore, the Naka-Rushton model better fitted the data from the prior experiment.

Task 1



Supplementary Figure 1. Comparison between models. $\Delta AICc$ (y-axis) represents the difference of $AICc$ between the original and the variant (i.e., Gaussian, linear, Naka-Rushton) models (x-axis), such as $\Delta AICc = AIC_{original} - AIC_{variant}$. Higher values indicate better model fit, and values above 0 indicate that the variant outperformed the original model. The comparison between models was performed for the (a) prior, (b) reward, and (c) sensory uncertainty experiments. In (a), the data is averaged for each model and group, across base rate, and the sample size consisted of 30 autistic and 42 non-autistic participants. In (b), the data is averaged for each model and group, across reward block, and the sample size consisted of 27 autistic and 42 non-autistic participants. In (c), the data is averaged for each model, and the sample size consisted of 27 autistic and 40 non-autistic participants.

Reward experiment

The ANOVA revealed a main effect of model ($F(1.53, 102.56) = 51.63, p < .001, \eta_p^2 = .44$), where the Naka-Rushton model was a better predictor than both the linear ($t(69) = 8.24, p < .001$) and Gaussian ($t(69) = 6.94, p < .001$) models. The main effect of reward was significant ($F(1, 67) = 42.05, p < .001, \eta_p^2 = .39$), with models overall fitting better on the data from the balanced, compared to the balanced block. The main effect of group was not significant ($F(1, 67) = 0.76, p = .386, \eta_p^2 = .01$), but the interaction between group and model was significant ($F(1.53, 102.56) = 3.69, p = .039, \eta_p^2 = .05$), and caused by a better fit of the Gaussian ($t(67) = 2.44, p = .018$) and the Naka-Rushton ($t(67) = 2.04, p = .045$) models for the autistic, compared to the non-autistic group. The difference between groups was not significant in the linear model ($t(67) = 0.92, p = .356$). The interaction between model and reward block was significant ($F(1.76, 117.6) = 19.59, p < .001, \eta_p^2 = .23$), and caused by a larger difference of $\Delta AICc$ between reward block in the linear ($t(68) = 6.64, p < .001$) and Gaussian ($t(68) = 4.03, p < .001$), compared to the Naka-Rushton ($t(68) = 2.13, p = .037$) model. The interaction between group and reward block ($F(1, 67) = 0.62, p = .436, \eta_p^2 < .01$) and the three-way interaction ($F(1.76, 117.6) = 1.44, p = .242, \eta_p^2 = .02$) were not significant. The one-sample t-test performed on $\Delta AICc$ of the Naka-Rushton model was significantly greater than 0 ($t(68) = 2.39, p = .020$), indicating that this model provided a better fit to the data than the original one. Therefore, the Naka-Rushton model better fitted the data from the reward experiment and outperformed the original model.

Sensory uncertainty experiment

The ANOVA performed on $\Delta A/C_c$ revealed no significant main effects of model ($F(1,37, 81.98) = 1.71, p = .194, \eta_p^2 = .03$), group ($F(1, 60) = 0.01, p = .922, \eta_p^2 < .01$), and no significant interaction between model and group ($F(1.37, 81.98) = 0.67, p = .463, \eta_p^2 = .01$).

These results indicate that the model estimating sensitivity using a Naka-Rushton function provided the best fit to the data in Experiments 1 and 2. In Experiment 3, all variants of the model performed similarly; however, for consistency, we analysed the Naka–Rushton fitted parameters in all subsequent analyses. After selecting this variant, we performed model comparisons between the model with meta-uncertainty as a free parameter (meta-uncertainty mode) and the identical with meta-uncertainty fixed at 0 (restricted model), revealed that, for Experiments 1 and 2, the model with the meta-uncertainty component outperformed the restricted model ($t(70)=9.35, p < .001$; $t(67) = 8.43, p < .001$, respectively), and this for both group ($p < .001$ for each group). However, in Experiment 3, the model with meta-uncertainty outperformed the restricted model for the non-autistic group ($t(32) = 3.50, p = .001$), but not the autistic group ($t(21) = 0.15, p = .880$). This indicates that, in the sensory uncertainty experiment, meta-uncertainty does not significantly explain the variance in the confidence behaviour in the autistic participants, suggesting a lower meta-uncertainty in the autistic group. The remaining analysis focuses on the fitted parameters of the meta-uncertainty model with the Naka-Rushton function to estimate sensitivity.

Category report

Prior experiment

The linear mixed-effect model performed on the proportion of reporting Category B revealed a significant main effect of orientation ($t(197.40) = 28.24, p < .001$), indicating that the proportion of reporting B increased as stimulus orientation became more clockwise. The main effect of contrast was not significant ($t(325.20) = 1.42, p = .156$), but the significant interaction between contrast and orientation ($t(10520) = 22.18, p < .001$) indicated a greater effect of orientation as contrast increased. Finally, the three-way interaction between orientation, contrast, and base rate was significant ($t(10510) = 3.25, p = .001$), indicating that the interaction between orientation and contrast was stronger when the base rate was unbalanced. All remaining main effects and interactions were not significant, including the main effect of group ($t(149.30) = 0.13, p = .898$) and all interactions associated with this factor (see **Supplementary Table 1**).

Reward experiment

The main effect of orientation on the probability of report category B was significant ($t(225.70) = 30.85, p < .001$), as well as the interaction between contrast and orientation ($t(10170) = -21.80, p < .001$). All remaining effects were not significant, including the main

effect of group ($t(144.90) = 0.04, p = .969$) and all interactions with this factor (see **Supplementary Table 1**).

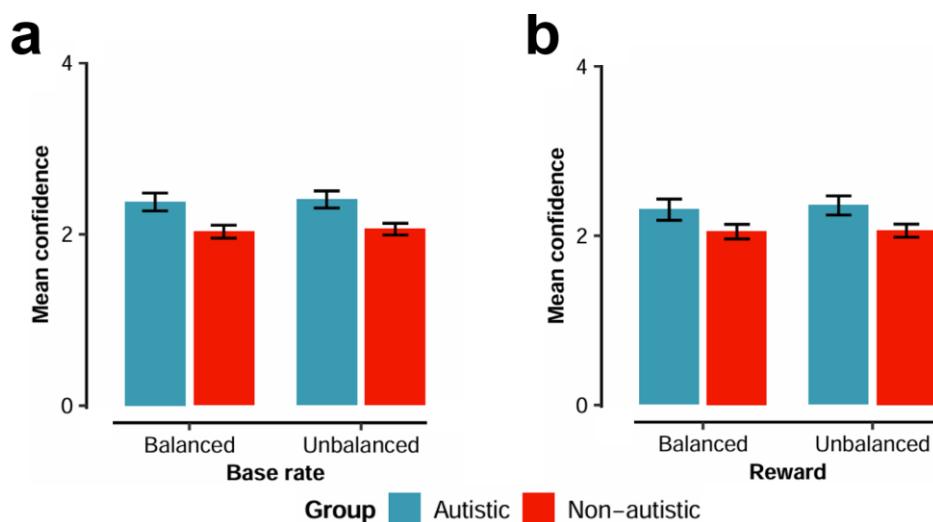
Sensory uncertainty experiment

The quadratic model performed on the proportion of reporting B revealed a main effect of contrast ($t(46.57) = 4.41, p < .001$), indicating that the proportion of reporting B increased with higher contrasts. The main effect of squared orientation was significant ($t(114.00) = 25.79, p < .001$), demonstrating the 'V' shape of the category report, with an increase of proportion of reporting B as orientations deviate from 0° . The interaction between the squared orientation and contrast was significant ($t(4528) = -18.96, p < .001$), demonstrating that the proportion of reporting B flattened as contrast decreased. All other effects were not significant (see **Supplementary Table 1**).

These results indicate that the category report was more sensitive to orientation as contrast increased, without differences between groups, in all three experiments.

Confidence report

The results from the linear and quadratic mixed-effect models investigating the confidence report in each experiment are reported in **Supplementary Table 2**. The main effects of base rate and reward conditions on confidence reports are illustrated in **Supplementary Figure 2**.

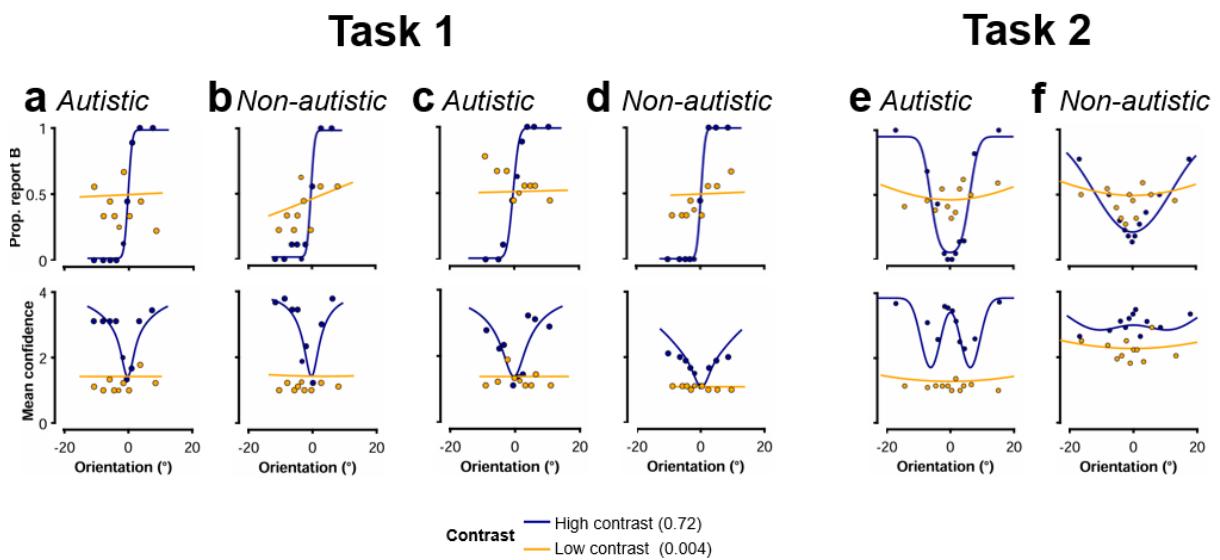


Supplementary Figure 2. Effect base rate/reward on the confidence report.

Mean confidence report (y-axis) for each base rate/reward condition (x-axis) and group (bar colour), for (a) the prior and (b) reward experiments. The data is averaged across orientations and contrasts. Bars show means across participants, and error bars represent \pm SE. The sample size consisted of 30 autistic and 41 non-autistic participants in (a) and 27 autistic and 41 non-autistic participants in (b).

Consistency in category and confidence report

Supplementary Figure 3 illustrates how category and confidence reports were predicted by the meta-uncertainty model. The figure displays proportion of reporting Category B (**top row**) and mean of confidence report (**bottom row**) as a function of stimulus orientation and the two extreme contrast values, for the prior (**a, b**), reward (**c, d**), and sensory evidence experiments (**e, f**). Here, we plotted observed and model predicted data for an individual subject in each subplot. The sigmoid, 'V', and 'W' shapes observed in **Fig. 2** for high contrasts—characteristic of category and confidence reports that are sensitive to stimulus information—were reproduced by the model predictions, and fitted properly with the observed data. For low contrasts, the fitted lines flattened, reproducing well the reduced association of category and confidence reports with stimulus information. These observations demonstrate that the pattern of category and confidence reports was well captured by the meta-uncertainty model. Importantly, in the sensory uncertainty experiment, we noticed that autistic individuals exhibited a more detailed association between reports and stimulus information, as illustrated by the steeper curve in **Task 2** for the autistic participant, suggesting a greater association between confidence and choice consistency. Therefore, participants' behaviour was well captured by the meta-uncertainty in all experiments.



Supplementary Figure 3. Model fitting for a sampled participant from each group in each experiment. Proportion of reporting Category B (**top row**, y-axis) and mean of confidence report (**bottom row**, y-axis) as a function of stimulus orientation (x-axis) and the two extreme contrast values, for the prior (**a, b**), reward (**c, d**), and sensory evidence experiments (**e, f**). Each subplot displays the observed and predicted behaviour for an individual participant. Data points illustrate choice behaviour, with size proportional to the number of trials. The solid lines represent the fit of the meta-uncertainty model using the maximum likelihood estimation method. The model was fit to all data simultaneously for each experiment and group.

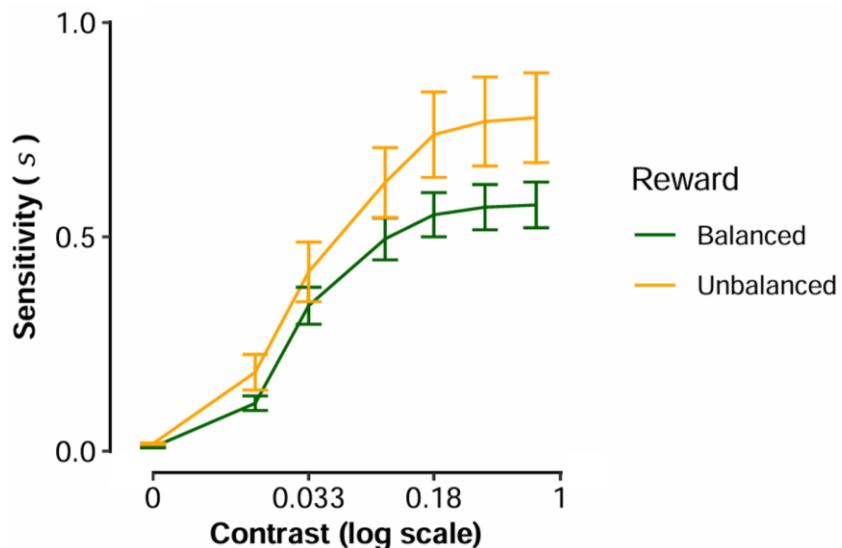
Perceptual sensitivity

Prior experiment

The ANOVA performed on the sensitivity (d') revealed no significant main effects of group ($F(1, 70) = 0.77, p = .782, \eta_p^2 < .01$), base rate ($F(1, 70) = 1.67, p = .201, \eta_p^2 = .02$), and no significant interactions between group and base rate ($F(1, 70) = 0.23, p = .630, \eta_p^2 < .01$), group and contrast ($F(1.52, 106.28) = 1.06, p = .336, \eta_p^2 = .02$), base-rate and contrast ($F(1.44, 101.01) = 0.51, p = .545, \eta_p^2 < .01$), and between contrast, group and base rate ($F(1.44, 101.01) = 0.28, p = .684, \eta_p^2 < .01$).

Reward experiment

The ANOVA performed on the sensitivity revealed no significant interactions between group and reward ($F(1, 67) = 2.01, p = .161, \eta_p^2 = .03$), group and contrast ($F(1.53, 102.80) = 3.04, p = .066, \eta_p^2 = .04$), and between group, reward and contrast ($F(1.31, 87.63) = 1.86, p = .174, \eta_p^2 = .03$). The main effect of reward was significant ($F(1, 67) = 8.83, p = .004, \eta_p^2 = .12$) and is illustrated in **Supplementary Figure 4**, reporting sensitivity as a function of contrast and reward block, across groups.



Supplementary Figure 4. Difference in sensitivity between reward conditions. Sensitivity as a function of contrast (x-axis) and reward block (line color). Data points and bars show means across participants and groups, and error bars represent \pm SE. The sample size consisted of 27 autistic and 41 non-autistic participants.

Decision criterion

Prior experiment

The ANOVA performed on the decision criterion revealed no significant interactions between group and base rate ($F(1, 70) = 2.11, p = .151, \eta_p^2 = .03$), group and contrast ($F(1.97, 138.18) = 2.62, p = .077, \eta_p^2 = .04$), and between group, base rate and contrast ($F(1.97, 138.18) = 2.62, p = .077, \eta_p^2 = .04$).

Reward experiment

The interaction between group and reward ($F(1, 69) = 0.11, p = .738, \eta_p^2 < .01$), group and contrast ($F(1.43, 98.42) = 0.07, p = .868, \eta_p^2 < .01$), and the triple interaction ($F(1.43, 98.43) = 0.07, p = .868, \eta_p^2 < .01$), were not significant.

Confidence criterion

Prior experiment

The interaction between group and base rate ($F(1, 70) = 0.10, p = .758, \eta_p^2 < .01$), group and confidence ($F(1.52, 106.47) = 0.22, p = .743, \eta_p^2 < .01$), base rate and confidence ($F(1.62, 113.53) = 2.154, p = .130, \eta_p^2 = .03$), and the triple interaction ($F(1.62, 113.53) = 2.53, p = .095, \eta_p^2 = .04$) were not significant.

Reward experiment

The interactions between group and reward ($F(1, 67) = 0.95, p = .335, \eta_p^2 = .01$), group and confidence ($F(1.56, 104.32) = 0.47, p = .581, \eta_p^2 < .01$), reward and confidence ($F(2, 134) = 2.10, p = .126, \eta_p^2 = .03$), and the three-way interaction ($F(2, 134) = 1.61, p = .203, \eta_p^2 = .02$) were not significant.

Guess rate

Prior experiment

The model integrated a measure of guess rate (g) to account for random reporting. The ANOVA performed on g showed a main effect of group ($F(1, 70) = 5.93, p = .017, \eta_p^2 = .08$), with a higher g for the autistic group, $t(50.8) = 2.32, p = .024$. The main effect of base rate ($F(1, 70) = 0.04, p = .850, \eta_p^2 < .01$), and the interaction between base rate and group ($F(1, 70) = 0.33, p = .569, \eta_p^2 < .01$) were not significant.

Reward experiment

The ANOVA on g revealed that g was not significantly different across groups ($F(1, 67) = 0.11, p = .746, \eta_p^2 < .01$) and reward ($F(1, 67) = 1.27, p = .264, \eta_p^2 = .02$), and the

interaction between group and reward was not significant ($F(1, 67) = 0.43, p = .517, \eta_p^2 < .01$).

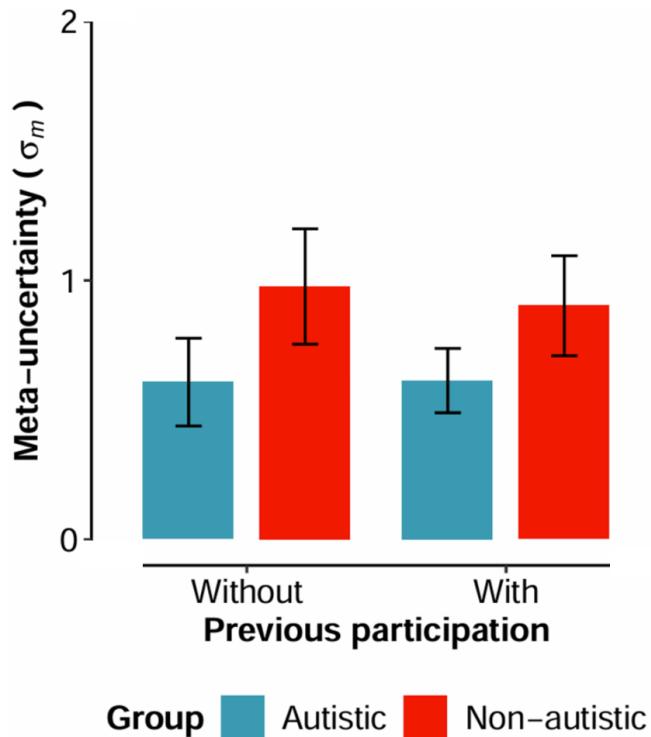
Sensory uncertainty experiment

The unpaired t-test investigating the difference in g between groups was not significant ($t(55.81) = 0.84, p = .403$).

Differences in meta-uncertainty within and between groups in the sensory uncertainty experiment (Experiment 3), based on familiarity with the task

Supplementary Figure 5 illustrates meta-uncertainty in Experiment 3 (sensory uncertainty manipulation) as a function of previous participation in other experiments and group. Six autistic and 16 non-autistic participants did not participate in the prior or reward experiments before completing the sensory uncertainty experiment. Sixteen autistic and 18 non-autistic participants completed at least one experiment prior to the sensory uncertainty experiment.

The ANOVA performed on meta-uncertainty revealed no main effects of previous participation ($F(1, 52) = 0.03, p = .875, \eta_p^2 < .01$), nor interaction between previous participation and group ($F(1, 52) = 0.03, p = .854, \eta_p^2 < .01$) (see **Supplementary Figure 5**). These results indicate that participants who completed Experiments 1 and 2 before completing Experiment 3 do not exhibit different metacognitive abilities compared to participants who performed the task for the first time. As most of the autistic participants—compared to non-autistics—performed Experiments 1 and 2 before completing Experiment 3, these results indicate that enhanced metacognitive abilities in the autistic group cannot be associated with training of confidence abilities. The main effect of group was not significant ($F(1, 52) = 2.35, p = .133, \eta_p^2 = .05$).



Supplementary Figure 5. Meta-uncertainty in Experiment 3 (sensory uncertainty manipulation) based on familiarity with the task. Meta-uncertainty (y-axis) as a function of previous participation—whether participants completed Experiments 1 or 2 (i.e., with) before participating in Experiment 3—and group (bar colour). Bars show means across participants, and error bars represent \pm SE. The sample size consisted of 22 autistic and 34 non-autistic participants.

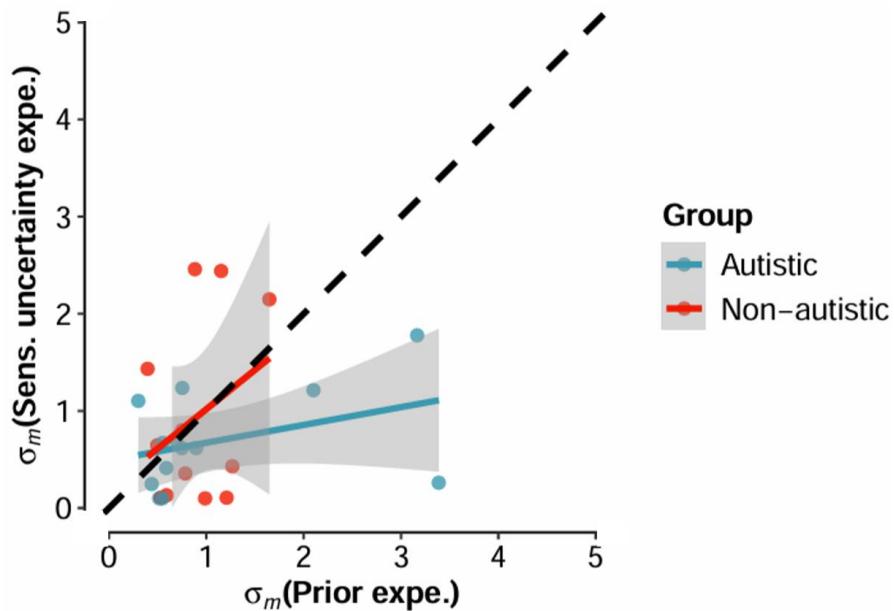
Within-subject differences in meta-uncertainty across sensory uncertainty and prior experiments

The main analyses of meta-uncertainty showed that non-autistic participants exhibited consistent meta-uncertainty across experiments, whereas autistic participants demonstrated lower meta-uncertainty (i.e., enhanced metacognition) when the first-order decision integrated sensory uncertainty alone, and greater meta-uncertainty (i.e., reduced metacognition) when prior information was integrated into the inference process. To test whether this pattern appeared at an individual level, we examined how within-subject meta-uncertainty varied between the two experiments for each group. The sample included participants who completed both experiments (13 autistic and 12 non-autistic).

Supplementary Figure 6 illustrates meta-uncertainty in the sensory uncertainty experiment as a function of meta-uncertainty in the prior experiment, per group. Meta-uncertainty values in the prior experiment are averaged across base rate blocks for each participant.

Due to the small sample size, participants' behaviour was examined qualitatively by visualizing trends. Here, the trends aligned with the previous findings: the interaction suggests a steeper relation between the meta-uncertainty values in the two experiments for

the non-autistic group. In **Supplementary Figure 6**, the regression line for the non-autistic group follows the $y = x$ diagonal, indicating similar meta-uncertainty across experiments. For the autistic group, the line falls below the diagonal, reflecting higher meta-uncertainty in the prior compared to the sensory uncertainty experiment within the same participants. These observations support our previous findings, indicating that, unlike non-autistic participants, metacognitive performance in autistic participants depends on the first-order Bayesian source of uncertainty.



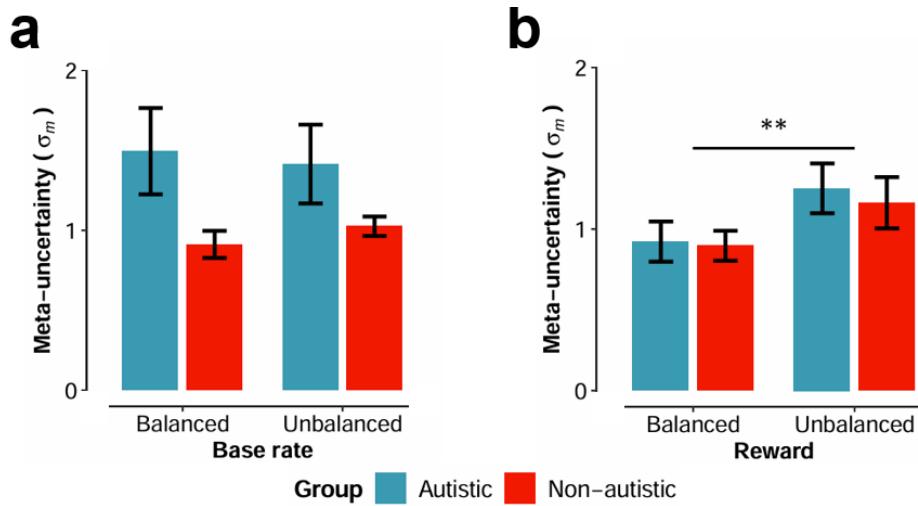
Supplementary Figure 6. Within-subject comparison of meta-uncertainty across the prior and the sensory uncertainty experiments. Meta-uncertainty from the sensory uncertainty experiment (y-axis) as a function of meta-uncertainty from the prior experiment (x-axis) per group (dot and line colour). Each dot represents meta-uncertainty values from both experiments for one participant. Regression lines (coloured lines) were fitted per group using linear models, and shaded areas represent the 95% confidence intervals around the regression lines. The sample size consisted of 13 autistic and 12 non-autistic participants

Meta-uncertainty analyses per group and block condition

Supplementary Figure 7 illustrates meta-uncertainty as a function of base rate/block condition and group.

Prior experiment

The ANOVA performed on the σ_m revealed a main effect of group ($F(1, 70) = 4.93, p = .030, \eta_p^2 = .07$), with the autistic group exhibiting a higher σ_m compared to the non-autistic group (see **Supplementary Figure 7**). The main effect of base rate ($F(1, 70) = 0.04, p = .848, \eta_p^2 < .01$) and the interaction between base rate and group ($F(1, 70) = 1.31, p = .256, \eta_p^2 = .02$) were not significant. Therefore, autistic individuals showed reduced metacognitive abilities across prior conditions.



Supplementary Figure 7. Analyses of meta-uncertainty per block condition. The meta-uncertainty σ_m (y-axis) as a function of block condition (x-axis) and group (bar colour) for the **(a)** prior and **(b)** reward experiments. Bars show means across participants, and error bars represent \pm SE. In **(b)**, the main effect of reward was evaluated using a mixed-design ANOVA. **.01 > $p \geq .001$. The sample size consisted of 30 autistic and 42 non-autistic participants **(a)** and 27 autistic and 42 non-autistic participants **(b)**.

Reward experiment

The ANOVA performed on σ_m revealed no significant difference between groups ($F(1, 67) = 0.10, p = .758, \eta_p^2 < .01$). The main effect of reward was significant ($F(1, 67) = 8.48, p = .005, \eta_p^2 = .11$), with a higher σ_m when reward was unbalanced compared to balanced, and the interaction between reward and group was not significant ($F(1, 67) = 0.10, p = .758, \eta_p^2 < .01$) (see **Supplementary Figure 7b**). Therefore, the autistic group exhibited similar metacognitive abilities compared to the non-autistic group when reward information was included in their perceptual decisions. Surprisingly, reward information influenced meta-uncertainty, and this in a similar manner between the two groups.

Supplementary Table

Supplementary Table 1. Result of the linear and quadratic mixed-effect models investigating category report for each experiment. The table reports the degree of freedom (*df*), *t*-value (*t*), and *p*-value (*p*) for each predictor (main effects and interactions). Orientation² indicates the squared predictor. The asterisks represent the significance levels, **p* < .05, ***p* < .01, ****p* < .001.

Prior experiment			
Predictor	df	t	p
Intercept	148.50	36.16	< .001***
Orientation	197.40	28.24	< .001***
Contrast	325.20	-1.42	.156
Group	149.30	-0.13	.898
Base rate	285.30	-0.25	.803
Orientation x Contrast	10520.00	-22.18	< .001***
Orientation x Group	197.90	0.79	.433
Contrast x Group	325.70	1.33	.186
Orientation x Base rate	10510.00	-0.44	.661
Contrast x Base rate	286.20	1.07	.284
Group x Base rate	10520.00	0.26	.792
Orientation x Contrast x Group	10510.00	1.03	.301
Orientation x Contrast x Base rate	10510.00	3.25	.001**
Orientation x Group x Base rate	10510.00	-0.71	.478
Contrast x Group x Base rate	10510.00	-0.78	.434
Orientation * Contrast * Group * Base rate	10510.00	0.22	.823
Reward experiment			
Predictor	df	t	p
Intercept	144.00	37.72	< .001***
Orientation	225.70	30.85	< .001***
Contrast	139.70	0.35	.725
Group	144.90	0.04	.969
Reward	1430.00	0.69	.494
Orientation x Contrast	10170.00	-21.80	< .001**
Orientation x Group	226.80	-1.06	.289
Contrast x Group	141.10	-0.80	.425
Orientation x Reward	10170.00	0.78	.433
Contrast x Reward	10160.00	-0.89	.375
Group x Reward	1438.00	0.01	.990

Orientation x Contrast x Group	10170.00	1.01	.311
Orientation x Contrast x Reward	10170.00	-0.34	.731
Orientation x Group x Reward	10170.00	-0.96	.339
Contrast x Group x Reward	10170.00	0.10	.920
Orientation x Contrast x Group x Reward	10170.00	1.01	.315
Sensory uncertainty experiment			
Predictor	df	t	p
Intercept	27.68	3.48	.002**
Orientation	177.40	-1.08	.281
Contrast	46.57	4.41	< .001***
Group	27.68	1.03	.313
Orientation ²	114.00	25.79	< .001***
Orientation x Contrast	4528.00	0.29	.776
Orientation x Group	177.40	2.62	.010**
Contrast x Group	46.56	-1.24	.222
Contrast x Orientation ²	4528.00	-18.96	< .001***
Group x Orientation ²	114.00	-1.60	.112
Group x Orientation x Contrast	4528.00	-1.15	.250
Group x Orientation ² x Contrast	4528.00	1.67	.096

Supplementary Table 2. Result of the quadratic mixed-effect models investigating confidence report for each experiment. The table reports the degree of freedom (df), t-value (t), and p-value (p) for each predictor (main effects and interactions). Orientation² indicates the squared predictor. The asterisks represent the significance levels, *p < .05, **p < .01, ***p < .001.

Prior experiment			
Predictor	df	t	p
Intercept	47.69	12.02	< .001
Orientation	39660.00	-0.01	.996
Contrast	29.12	5.70	< .001***
Group	47.70	-1.54	.130
Base rate	59960.00	2.82	.005**
Orientation ²	777.50	4.14	.770
Orientation x Difficulty	59970.00	-0.97	.333
Orientation x Group	39170.00	-0.01	.999
Contrast x Group	28.17	-1.38	.179
Orientation x Base rate	59970.00	0.72	.470
Contrast x Base rate	59960.00	-1.58	.114

Group x Base rate	59960.00	-1.70	.090
Contrast x Orientation ²	59960.00	6.91	< .001
Group x Orientation ²	781.20	0.80	.872
Base rate x Orientation ²	59960.00	-0.28	.783
Orientation x Contrast x Group	59970.00	3.00	.003**
Orientation x Contrast x Base rate	59970.00	0.85	.394
Orientation x Group x Base rate	59970.00	-0.35	.727
Contrast x Group x Base rate	59960.00	1.21	.226
Contrast x Group x Orientation ²	59960.00	1.22	.224
Contrast x Base rate x Orientation ²	59960.00	0.03	.979
Group x Base rate x Orientation ²	59970.00	1.28	.200
Orientation x Contrast x Group x Base rate	59970.00	-1.78	.075
Orientation ² x Contrast x Group x Base rate x	59960.00	-0.13	.894
Reward experiment			
Predictor	df	t	p
Intercept	25.54	10.17	.398
Orientation	6440.00	-0.02	.981
Contrast	3.99	4.75	.009**
Group	25.54	-0.71	.763
Reward	54860.00	2.96	.003**
Orientation ²	30.42	0.69	.746
Orientation x Difficulty	54860.00	0.74	.457
Orientation x Group	6369.00	0.02	.988
Contrast x Group	3.90	-1.84	.139
Orientation x Reward	54830.00	2.67	.008**
Contrast x Reward	55090.00	0.32	.751
Group x Reward	54850.00	-0.84	.402
Contrast x Orientation ²	51110.00	5.52	< .001***
Group x Orientation ²	30.41	-0.05	.977
Reward x Orientation ²	54880.00	-1.64	.101
Orientation x Contrast x Group	54890.00	-0.41	.685
Orientation x Contrast x Reward	54850.00	-2.03	.042*
Orientation x Group x Reward	54840.00	-2.03	.043*
Contrast x Group x Reward	55070.00	-0.46	.646
Contrast x Group x Orientation ²	55080.00	1.29	.198
Contrast x Reward x Orientation ²	55200.00	2.53	.012*
Group x Reward x Orientation ²	54880.00	0.05	.959

Orientation x Contrast x Group x Reward	54850.00	1.84	.065
Orientation ² x Contrast x Group x Reward	55190.00	-1.10	.269
Sensory uncertainty experiment			
Predictor	<i>df</i>	<i>t</i>	<i>p</i>
Intercept	15.42	11.84	< .001***
Orientation	22460.00	-0.01	.995
Contrast	5.01	5.65	.002**
Group	15.42	-0.10	.923
Orientation ²	801.20	0.25	.941
Orientation x Contrast	98760.00	-0.74	.457
Orientation x Group	22610.00	0.01	.993
Contrast x Group	5.00	-1.73	.145
Contrast x Orientation ²	98760.00	-1.54	.125
Group x Orientation ²	801.20	-0.01	.997
Group x Orientation x Contrast	98760.00	-0.14	.886
Group x Orientation ² x Contrast	98770.00	2.99	.003**